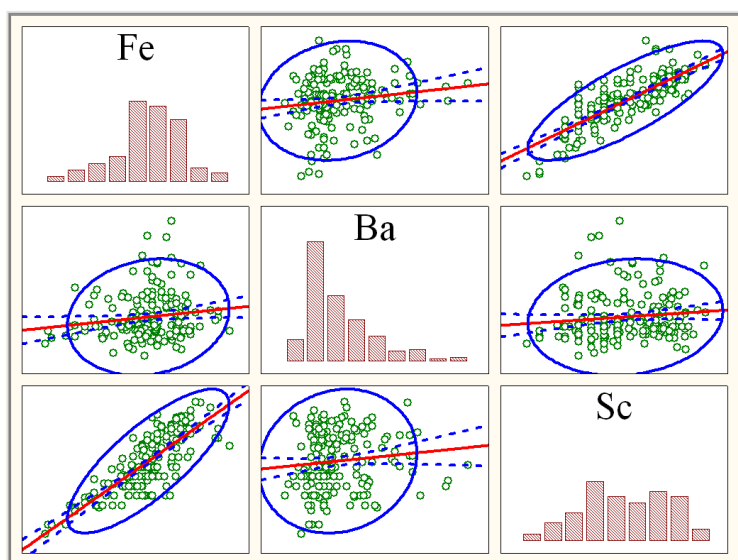


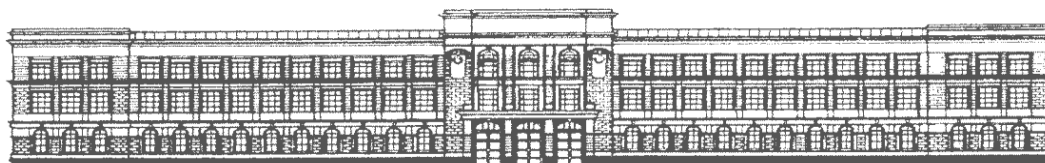


А.А. Михальчук, Е.Г. Язиков

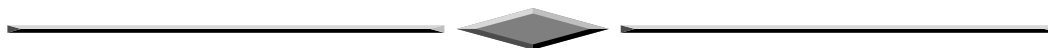
МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ



ЭКОЛОГО-ГЕОХИМИЧЕСКИХ ИЗМЕРЕНИЙ



**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**



А.А. Михальчук, Е.Г. Язиков

**МНОГОМЕРНЫЙ
СТАТИСТИЧЕСКИЙ АНАЛИЗ
ЭКОЛОГО-ГЕОХИМИЧЕСКИХ
ИЗМЕРЕНИЙ**

ЧАСТЬ III. ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Издательство ТПУ
Томск 2014

УДК 681.3 : 519.2

А.А. Михальчук, Е.Г. Язиков

Многомерный статистический анализ эколого-геохимических измерений. Часть III. Лабораторный практикум
Учебное пособие. - Томск: Изд. ТПУ, 2014.- 198 с.

Настоящее учебное пособие предназначено для студентов и аспирантов по специальности 320300 (013600)- «Геоэкология» и может быть полезным при усвоении теоретического материала и овладении необходимых практических навыков при проведении сравнительного статистического анализа (ССА) эколого-геохимической информации. Пособие содержит:

- * краткие теоретические сведения;
- * примеры и рекомендации по решению типовых задач с использованием современного компьютерного инструментария (систем STATISTICA и Excel) на уровне модульного анализа данных с помощью мастер - макросов;
- * учет особенностей ССА в случае малых выборок;
- * наглядные графические иллюстрации, выполненные в системе STATISTICA 6.1;
- * список рекомендуемой литературы.

УДК 681.3 : 519.2

Рекомендовано к печати Редакционно-издательским советом
Томского политехнического университета

Рецензенты

Д.ф.-м.н., профессор, зав. каф. ТФ ТГУ

А.В. Шаповалов

К.г.-м.н., зам. нач. экспед. ГТИиГК ООО «ВостокГАЗПРОМгеофизика»

О.А.Миков

© Томский политехнический университет, 2014

© А.А. Михальчук, Е.Г. Язиков, 2014

Оглавление

ЧАСТЬ III. ЛАБОРАТОРНЫЙ ПРАКТИКУМ	5
3.1. Статистический анализ эколого-геохимических оценок разных территорий	5
3.1.1. Вычисление выборочных характеристик содержаний химических элементов в почве территорий.....	5
3.1.2. Проверка гипотезы о нормальном распределении содержаний микроэлементов в почве территорий.....	10
3.1.3. Сравнение содержаний химических элементов по двум территориям.....	14
3.1.4. Корреляционный анализ содержаний микроэлементов в почве территорий.....	23
3.1.5. Регрессионный анализ содержаний микроэлементов в почве территорий.....	32
3.1.6. Кластерный анализ содержаний микроэлементов в почве территорий.....	38
3.1.7. Канонический анализ двух ассоциаций микроэлементов в почве территорий.....	42
3.2. Статистический анализ эколого-геохимической информации по данным различных съемок	55
3.2.1. Вычисление выборочных характеристик содержаний химических элементов в ассоциациях по трем средам.....	55
3.2.2. Однофакторный дисперсионный анализ содержаний химических элементов в ассоциациях по трем средам.....	59
3.2.3. Корреляционный анализ содержаний микроэлементов в ассоциациях по трем средам.....	73
3.2.4. Регрессионный анализ содержаний микроэлементов в ассоциациях по трем средам.....	80
3.2.5. Кластерный анализ содержаний микроэлементов в ассоциациях по трем средам.....	83
3.2.6. Факторный анализ содержаний микроэлементов в снеговом осадке.....	88
3.2.7. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов.....	100

3.3. Статистический анализ содержаний микроэлементов в накипи разных территорий, полученных методами ИНАА и ISP.....	115
3.3.1. Парный сравнительный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.....	115
3.3.2. Двухфакторный дисперсионный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.....	123
3.3.3. Однофакторный дисперсионный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.....	129
3.3.4. Корреляционно-регрессионный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.....	136
3.1.5. Кластерный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.	141
3.3.6. Канонический анализ двух ассоциаций микроэлементов в накипи территории Ч, полученных методами ИНАА и ISP.	142
3.2.7. Факторный анализ содержаний микроэлементов в накипи территории Т, полученных методом ISP (выборка Т*І).....	149
3.3.8. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов в Т*І.....	158
3.3.9. Факторный анализ содержаний микроэлементов в накипи территории Ч, полученных методом ISP (выборка Ч*І).....	171
3.3.10. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов в Ч*І.....	178
ПРИЛОЖЕНИЯ.....	187
П ₂ . Содержание микроэлементов в почве территорий гг. Междуреченска и Томска, полученное методом ЭСП.....	187
П ₃ . Содержание микроэлементов в почве, снеге и золе растений территории г. Стрежевого.....	194
П ₄ . Категоризированное содержание микроэлементов в накипи территорий Томской и Челябинской областей, полученных методами ИНАА и ISP.....	197

ЧАСТЬ III. ЛАБОРАТОРНЫЙ ПРАКТИКУМ

3.1. Статистический анализ (СА) эколого-геохимических оценок разных территорий

Проведем СА содержаний химических элементов в почве территорий гг. Междуреченска (М) и Томска (Т), полученных методом ЭСП, поэтапно (из методических соображений аналогично п.2.1-4).

На предварительном этапе создается файл (с названием, напр., Т – М ЭСП почва.sta) категориальной и отдельно по территориям согласно баз данных (см. Приложение П₂). При этом названия химических элементов выносятся в названия переменных (столбцов).

3.1.1. Вычисление выборочных характеристик содержаний химических элементов в почве территорий (п. 2.1.1).

Шаг 1. Запустите в главном меню модуль **Анализ**, выберите

Основные статистики и таблицы, затем **Описательные статистики** и нажмите **ОК** (рис.3.1).

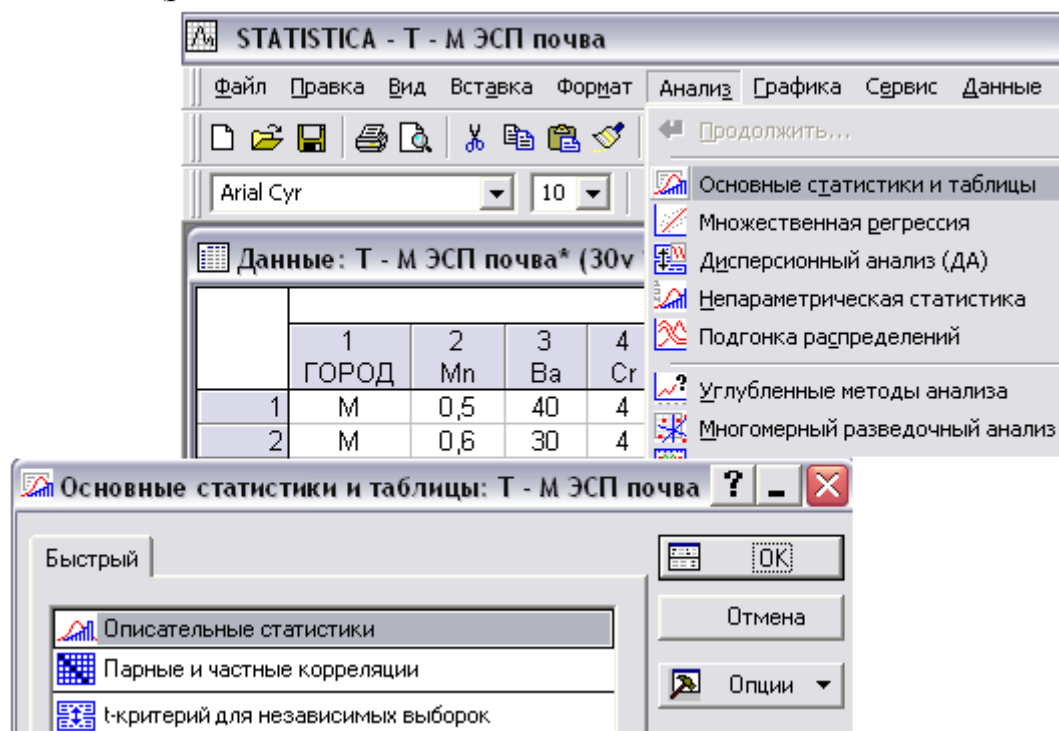


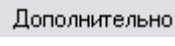

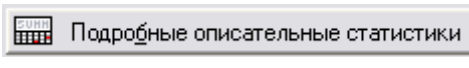


Рис.3.1. Открытие диалогового окна **Описательные статистики** в модуле **Анализ** системы STATISTICA

Шаг 2. В открывшемся окне нажмите кнопку  и выберите переменные базы данных М (рис.3.2) и нажмите .

Шаг 3. Нажмите кнопку  и закажите (поставьте все нужные числовые характеристики (рис.3.2)): объем выборки (N), среднее (m), медиана (Me), мода (Mo), стандартное отклонение (S), стандартная ошибка среднего (σ_m), асимметрию (A), эксцесс (E) и их стандартные ошибки (σ_A, σ_E), а также минимум (Min) и максимум (Max).

Шаг 4. Нажмите в окне  кнопку . Следующая таблица появится на экране (табл.3.1):

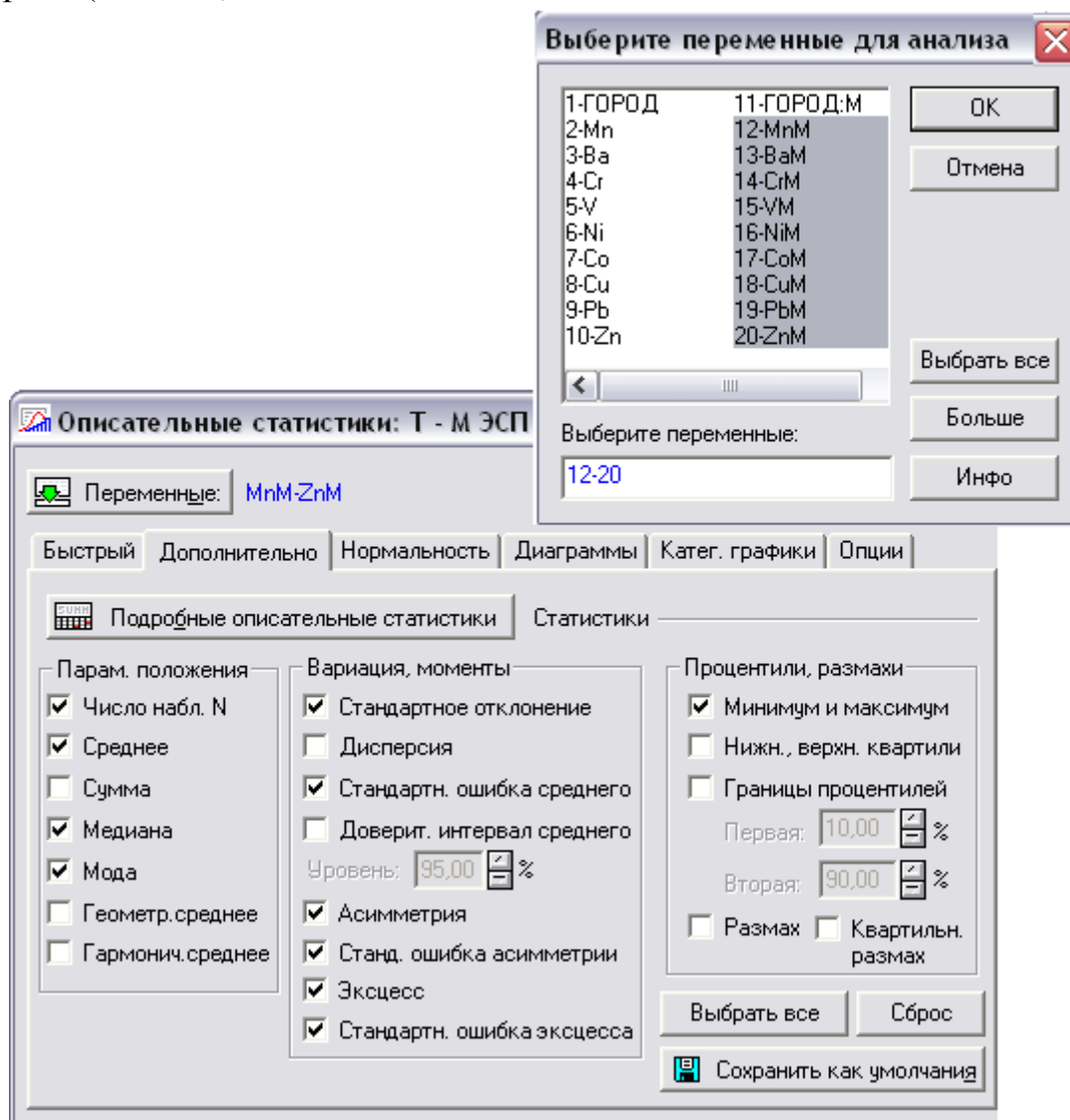


Рис.3.2. Режим  в окне .

Таблица 3.1

Оценки числовых характеристик содержаний микроэлементов в почво-грунтах М.

	<i>n</i>	<i>m</i>	<i>Me</i>	<i>Mo</i>	<i>Min</i>	<i>Max</i>	<i>S</i>	σ_m	<i>A</i>	σ_A	<i>E</i>	σ_E
Mn_М	100	0,39	0,40	0,40	0,10	0,80	0,20	0,02	0,30	0,24	-0,72	0,48
Ba_М	100	40,8	40,0	30,0	20,0	80,0	14,95	1,50	0,92	0,24	0,28	0,48
Cr_М	100	6,64	6,00	6,00	2,00	14,0	2,63	0,26	0,20	0,24	-0,37	0,48
V_М	100	5,65	6,00	6,00	1,00	10,0	2,10	0,21	-0,05	0,24	-0,45	0,48
Ni_М	100	3,52	3,00	3,00	1,00	7,00	1,12	0,11	0,65	0,24	1,01	0,48
Co_М	100	1,71	1,50		0,40	4,00	0,93	0,09	0,40	0,24	-0,67	0,48
Cu_М	100	5,03	5,00	4,00	2,00	10,0	1,80	0,18	0,58	0,24	0,17	0,48
Pb_М	100	3,35	3,00	4,00	1,00	6,00	1,27	0,13	0,07	0,24	-0,50	0,48
Zn_М	100	9,64	8,00	8,00	4,00	20,00	3,72	0,37	0,98	0,24	0,76	0,48

Аналогичным образом можно вычислить числовые характеристики содержаний микроэлементов в почво-грунте Т (табл.3.2).

Таблица 3.2

Оценки числовых характеристик содержаний микроэлементов в почво-грунтах Т.

	<i>n</i>	<i>m</i>	<i>Me</i>	<i>Mo</i>	<i>Min</i>	<i>Max</i>	<i>S</i>	σ_m	<i>A</i>	σ_A	<i>E</i>	σ_E
Mn_Т	100	0,42	0,40	0,40	0,10	0,80	0,19	0,02	0,13	0,24	-0,80	0,48
Ba_Т	100	40,4	40,0	30,0	20,0	80,0	15,6	1,56	0,83	0,24	-0,05	0,48
Cr_Т	100	11,02	10,0	10,0	2,00	20,0	4,47	0,45	0,24	0,24	-0,67	0,48
V_Т	100	6,29	6,00		1,00	10,0	2,21	0,22	-0,27	0,24	-0,44	0,48
Ni_Т	100	3,33	3,00	3,00	1,00	7,00	1,06	0,11	0,79	0,24	1,01	0,48
Co_Т	100	1,70	1,50	1,50	0,30	4,00	0,86	0,09	0,59	0,24	-0,09	0,48
Cu_Т	100	5,57	6,00	6,00	2,00	10,0	1,98	0,20	0,17	0,24	-0,50	0,48
Pb_Т	100	2,90	3,00	2,00	1,00	6,00	1,28	0,13	0,45	0,24	-0,29	0,48
Zn_Т	100	6,68	6,00	5,00	3,00	20,0	3,41	0,34	1,90	0,24	4,13	0,48

В табл. 3.1 отсутствует значение Mo для Co_M , что свидетельствует о мультимодальности распределения Co_M . Согласно табл.3.2 по аналогии с Co_M (табл.3.1), распределения V_T является мультимодальным. Наглядным подтверждением тому является гистограмма распределения, для построения которой необходимо выполнить последовательность следующих действий (п. 2.1.3):

Шаг 1. В модуле **Графика** выберите **Гистограммы...** (рис. 3.3).

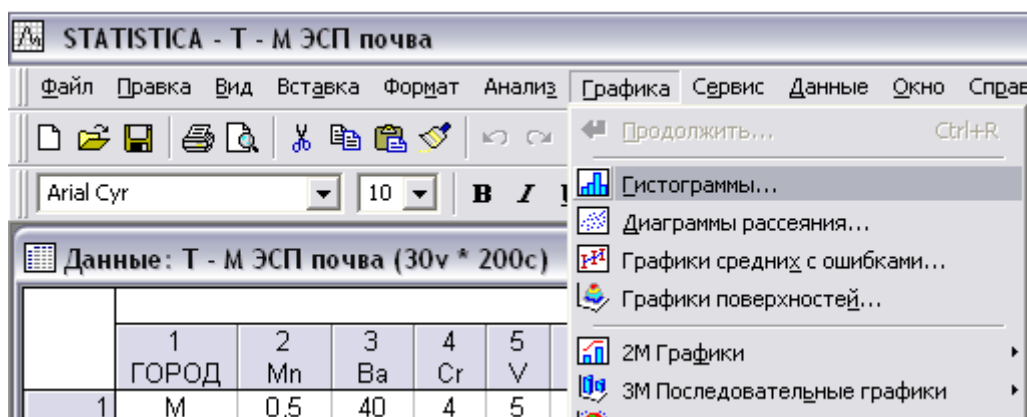


Рис.3.3. Запуск диалогового окна **Гистограммы...** в модуле **Графика**

Шаг 2. В режиме **Дополнительно** выберите параметры (см. рис.3.4).

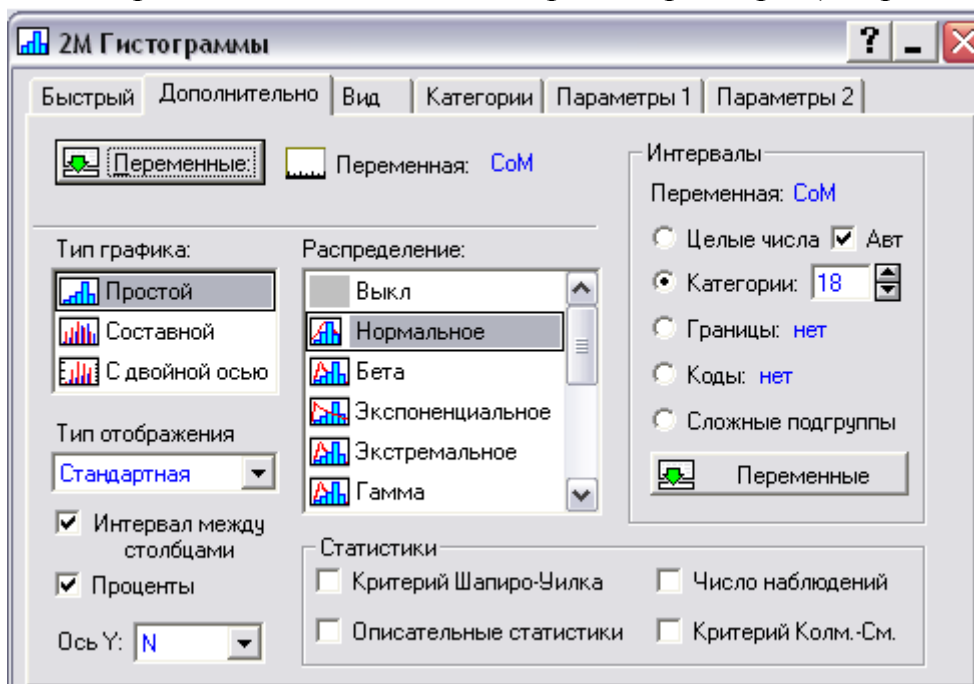


Рис. 3.4. Режим **Дополнительно** в окне **2M Гистограммы**
Следующая картинка появится на экране.

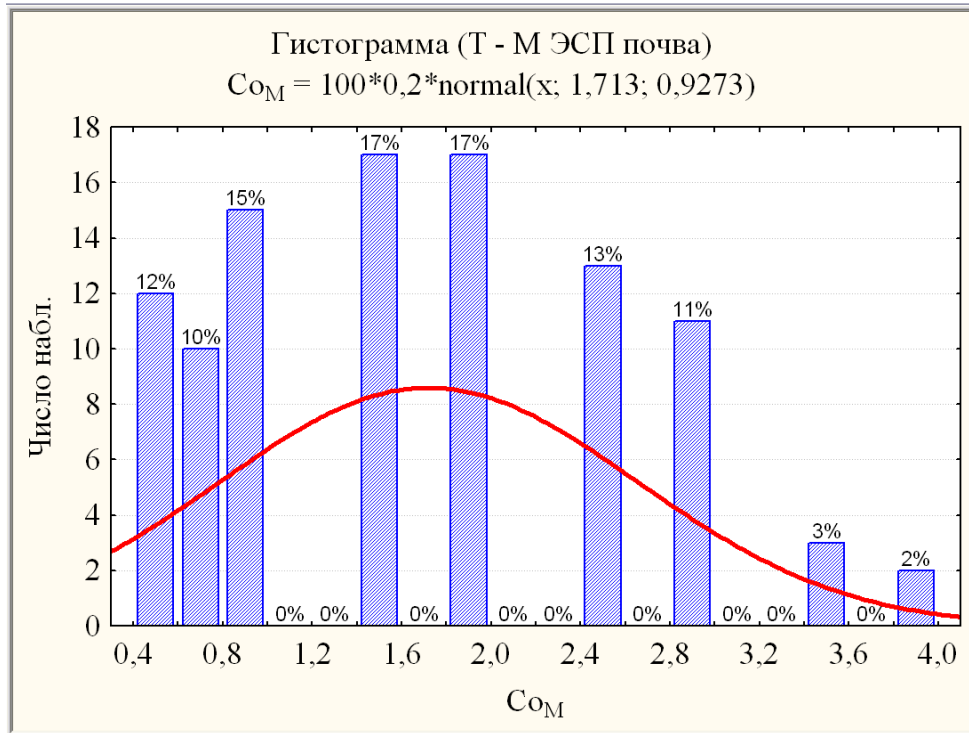


Рис. 3.5. Гистограмма распределения содержания C_{0M}

Аналогично строится гистограмма распределения содержания V_T

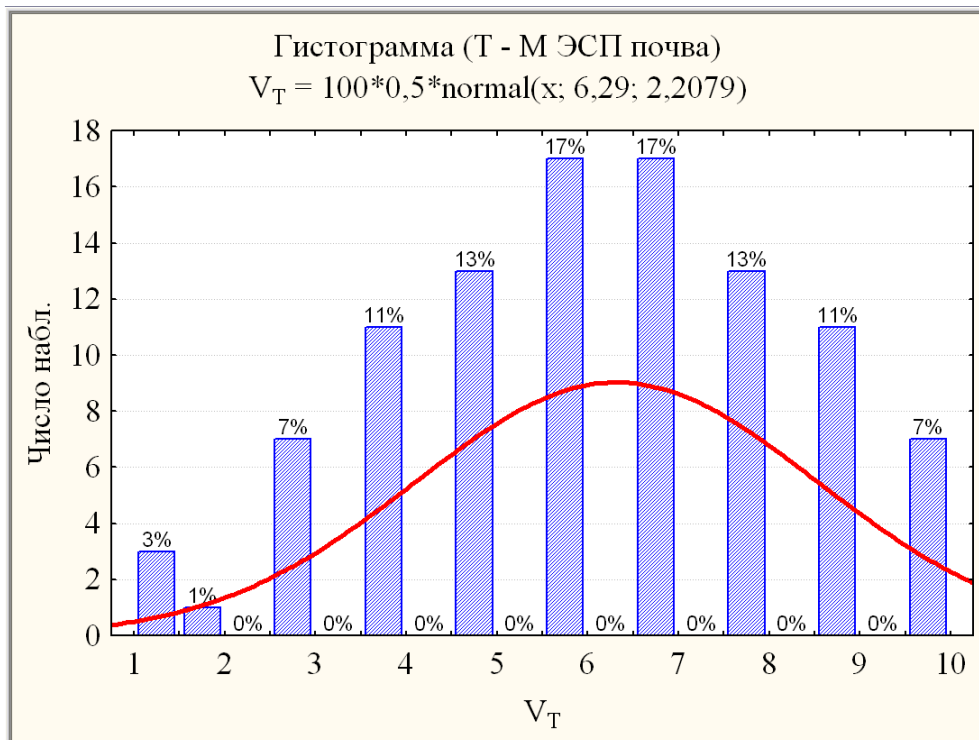


Рис. 3.6. Гистограмма распределения содержания V_T

3.1.2. Проверка гипотезы о нормальном распределении содержания микроэлементов в почве территорий (п. 2.2.1).

В связи с этим необходимо выполнить последовательность действий:

Шаг 1. Запустите модуль **Анализ** и нажмите **Подгонка распределений** (рис.3.7). В открывшемся окне **Подгонка распределений** выберите **Нормальное** распределение (рис. 3.7) нажмите **OK**.

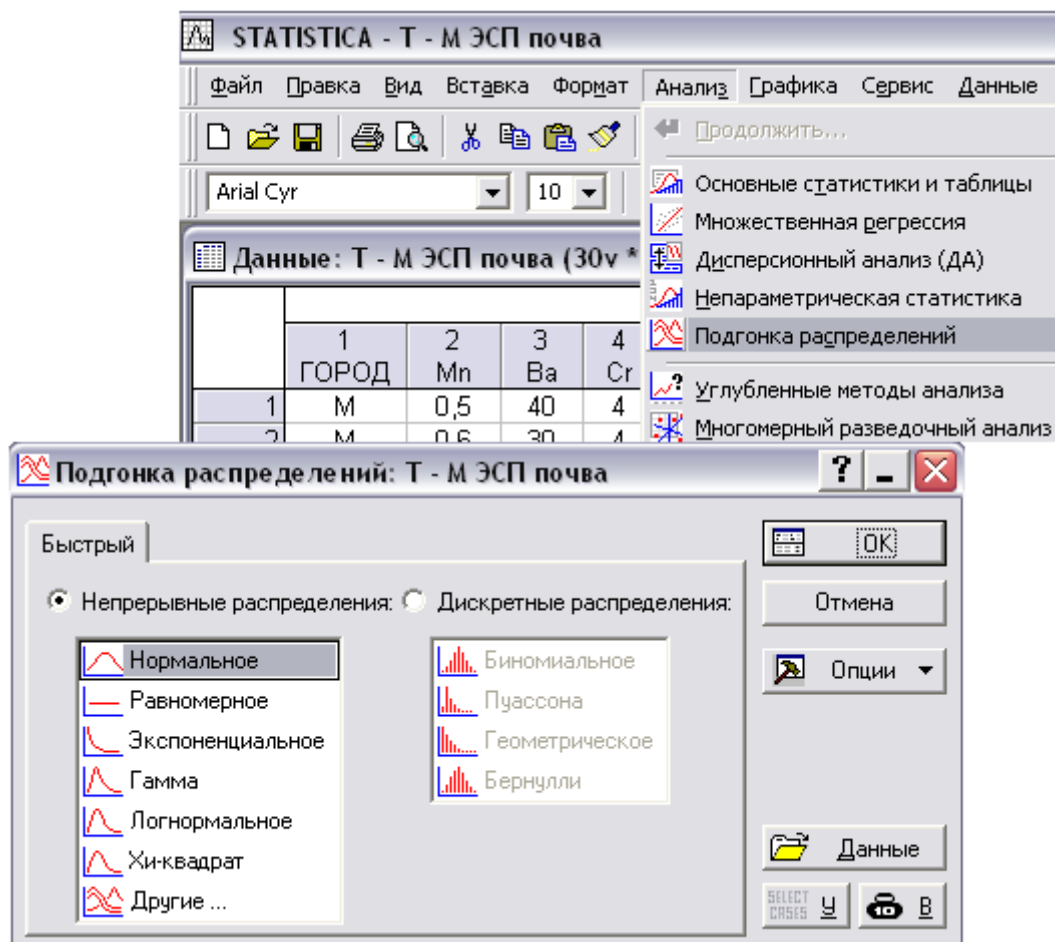


Рис.3.7. Окно **Подгонка распределений** в модуле **Анализ**

Шаг 2. В открывшемся окне нажмите кнопку **Переменная:**, выберите переменную Mn_M (рис.3.8) и нажмите **OK**.

Шаг 3. Нажмите кнопку **Опции** (рис.3.8) и закажите (поставьте точку) тест Колмогорова – Смирнова и (поставьте галочкой) тест Хи – квадрат, также нажмите кнопку **Параметры** (рис.3.8) и установите в окне

Число групп: число k столбцов гистограммы ($k=1+4lgn$, что дает $k = 9$ для объема выборки $n = 100$).

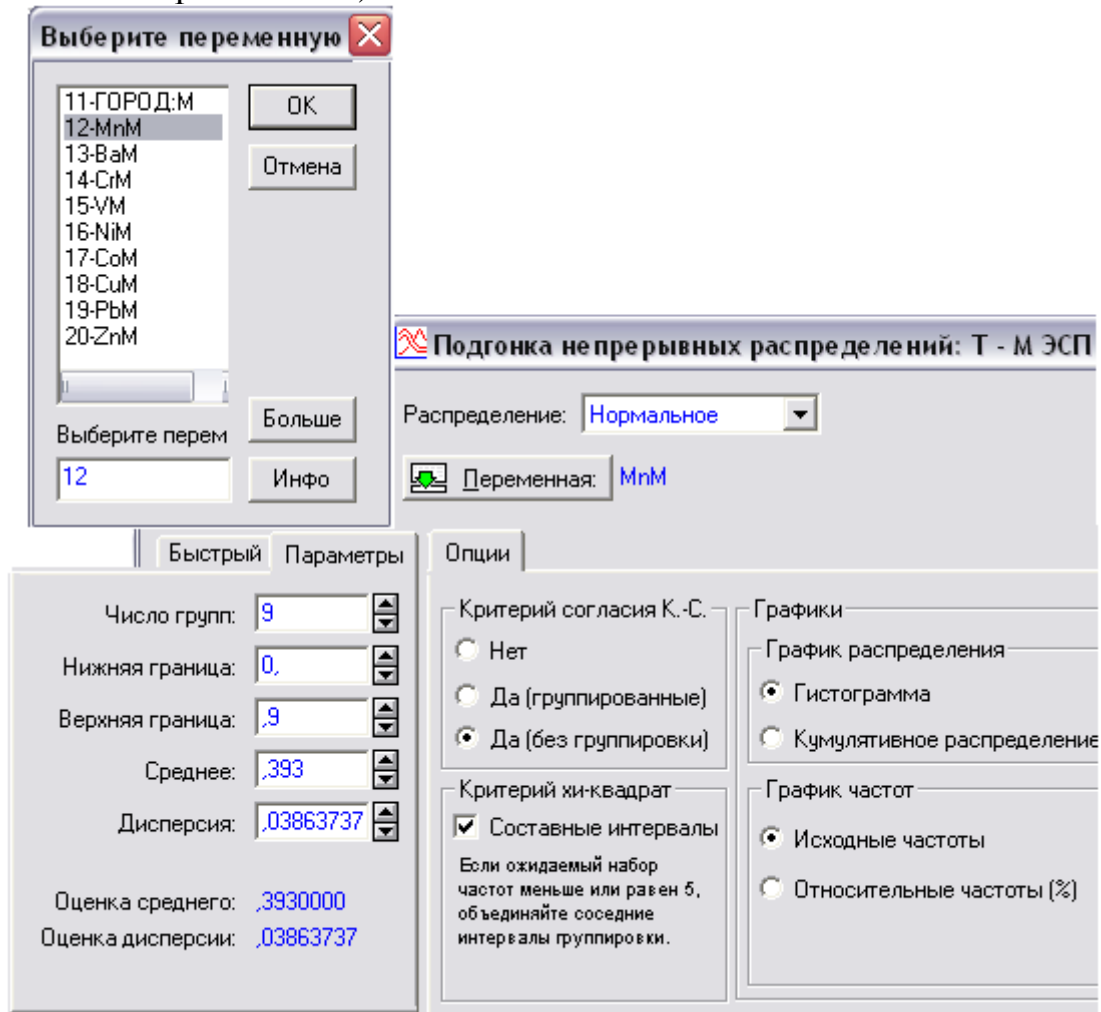
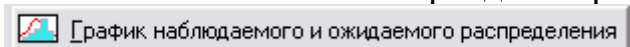


Рис.3.8. Окно **Подгонка не прерывных распределений: Т - М ЭСП почва**

Шаг 4. После чего перейдите в режим **Быстрый** и нажмите кнопку



На экране появится картинка для Mn_M (рис.3.9). В подзаголовке указаны уровни значимости p (тестов Колмогорова-Смирнова $p < 0,15$ и Хи-квадрат $p=0,06895$) отличия гистограммы от кривой плотности нормального распределения. В случае теста Колмогорова-Смирнова вместо неравенства можно вычислить соответствующий уровни значимости $p = \alpha_{K-S}$ по значению критерия d по формуле (1.21):

$$\alpha_{K-S} \approx 2 \exp \left[\sqrt{\frac{1}{2} + \left(1 + \frac{9}{2}n\right)^2} - 18n^2 \left(d + \frac{1}{6n}\right)^2 - \frac{9}{2}n - 1 \right] \text{ для } 0,01 < \alpha < 0,2 \text{ и } n > 10.$$

В данном случае получаем $p = \alpha_{к-с} \approx 0,094$. Подобные вычисления можно выполнить с помощью, например, системы Excel (рис.3.10), поместив значения n в ячейку A1 и d в ячейку B1, а в ячейке C1 запрограммировав формулу (1.21) с помощью строки формул.

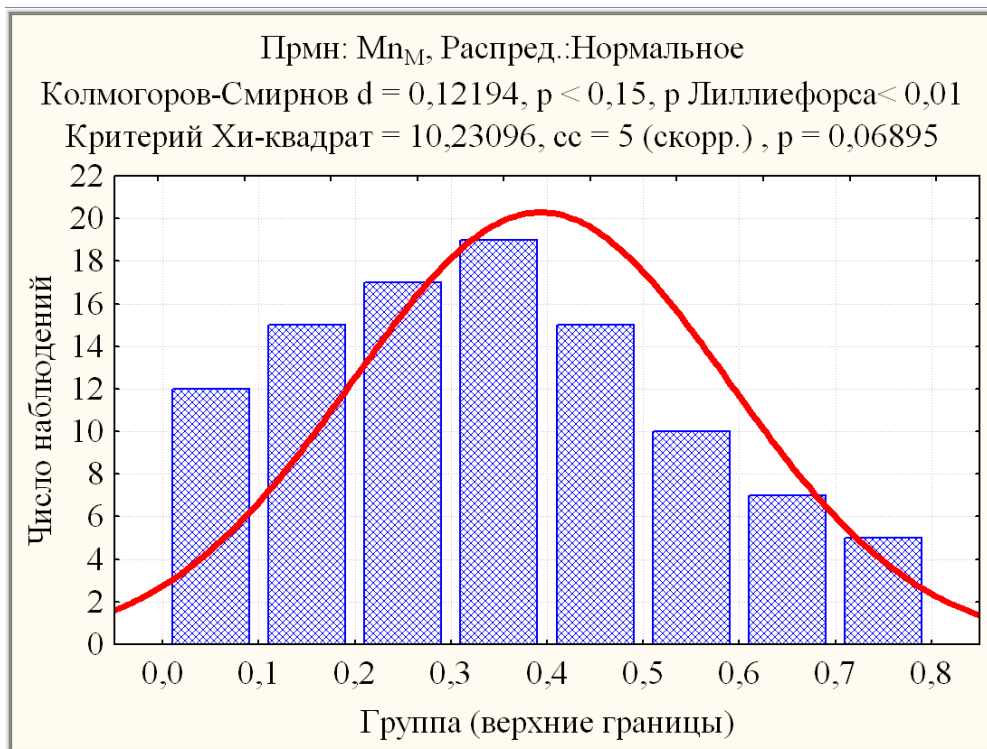


Рис. 3.9. Гистограмма распределения содержания Mn_M с кривой плотности нормального распределения и результатами тестов Колмогорова-Смирнова (1.21) и Хи-квадрат (1.19).

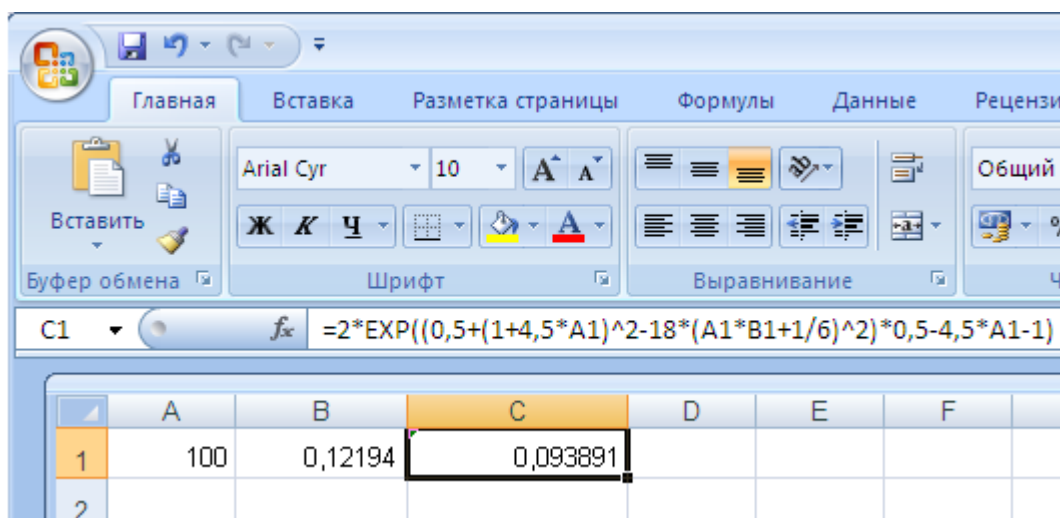


Рис. 3.10. Вычисление в Excel $\alpha_{к-с}$ (ячейка C1) с помощью формулы (1.21) по n (ячейка A1) и d (ячейка B1)

Степень значимости отличия экспериментального распределения от соответствующего нормального закона качественно определяется по уровню значимости (Боровиков, 2003): не значимые ($p \geq 0,100$), слабо значимые ($0,100 > p \geq 0,050$), статистически значимые ($0,050 > p \geq 0,005$), сильно значимые ($0,005 > p \geq 0,0005$), высоко значимые ($0,0005 > p$).

В качестве критерия соответствия эмпирического распределения теоретическому нормальному используют модули отношений (1.22) показателей асимметрии \tilde{A} и эксцесса \tilde{E} к их стандартным ошибкам $\sigma_{\tilde{A}}$ и $\sigma_{\tilde{E}}$ соответственно. Если эти отношения по абсолютной величине превышают 3, то гипотеза о нормальном распределении отвергается. Например, по табл.3.1 нет оснований отвергать гипотезу о нормальном распределении Mn_M . В табл.3.1 и 3.2 жирным шрифтом выделены случаи, когда гипотеза о нормальном распределении отвергается.

Таким образом, по совокупности трех критериев можно считать отличие эмпирического распределения Mn_M от нормального теоретического слабо значимым.

Аналогичным образом можно проверить гипотезу о нормальном распределении содержаний всех остальных химических элементов. Полученные таким образом результаты можно свести в одну таблицу 3.3.

Таблица 3.3

Критерии сравнения с нормальным законом распределения содержаний химических элементов в почво-грунтах М и Т

Элемент	М				Т			
	χ^2	α_{K-S}	\tilde{t}_1	\tilde{t}_2	α_{χ^2}	α_{K-S}	\tilde{t}_1	\tilde{t}_2
Mn	0,069	0,094	1,26	1,50	0,201	0,121	0,52	1,67
Ba	<10 ⁻³	<10 ⁻³	3,80	0,58	<10 ⁻³	<10 ⁻³	3,46	0,10
Cr	0,370	0,007	0,82	0,78	0,013	0,062	1,00	1,40
V	0,054	0,297	0,22	0,94	0,063	0,197	1,13	0,91
Ni	<10 ⁻³	0,017	2,71	2,10	<10 ⁻³	<10 ⁻³	3,29	2,10
Co	0,002	0,021	1,67	1,40	0,012	<10 ⁻³	2,46	0,19
Cu	0,386	0,024	2,40	0,36	0,067	0,214	0,72	1,04
Pb	<10 ⁻³	0,007	0,29	1,05	0,114	0,001	1,88	0,61
Zn	<10 ⁻³	<10 ⁻³	4,06	1,59	<10 ⁻³	<10 ⁻³	7,87	8,64

Согласно таблице 3.3 нормально распределенными содержаниями химических элементов можно считать Mn, V и Cu.






Критерии \tilde{t}_1 и \tilde{t}_2 рассчитываются по табл.3.1 - 2 с помощью калькулятора или в Excel.

Таблица 3.3 дает представление о разной степени значимости отличия эмпирических распределений содержаний химических элементов от нормального теоретического закона.

3.1.3. Сравнение содержаний химических элементов по двум территориям.

Применяем критерий Стьюдента (1.24-25) в случае нормально распределенными в обоих городах содержаниями химических элементов Mn, V, Cu и непараметрические критерии Манна-Уитни и Колмогорова-Смирнова (Боровиков В.П., 2003) в остальных случаях.

Для проведения оценки различия содержаний химических элементов по двум территориям посредством применения параметрического критерия Стьюдента необходимо выполнить последовательность следующих действий (п. 2.2.2):

Шаг 1. Нажмите  Основные статистики и таблицы в стартовой панели модуля  Анализ, выберите  t-критерий для независимых выборок (рис.3.11) и нажмите   OK.

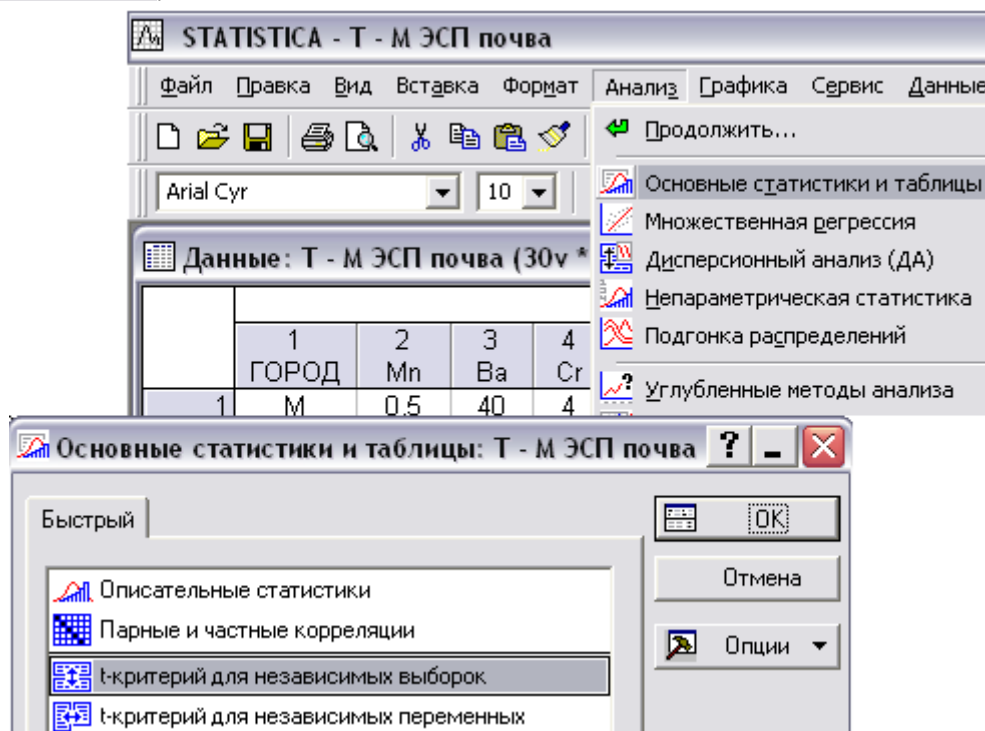


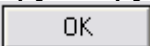


Рис.3.11. Открытие окна  t-критерий для независимых выборок в модуле  Анализ

Шаг 2. В открывшемся окне нажмите кнопку  и выберите переменные категориальной базы данных (переменная «Город» - группирующая, а микроэлементы Mn, V, Cu - зависимые) и нажмите . После этого в окнах кодов для групп появятся «М» и «Т» (рис.3.12).

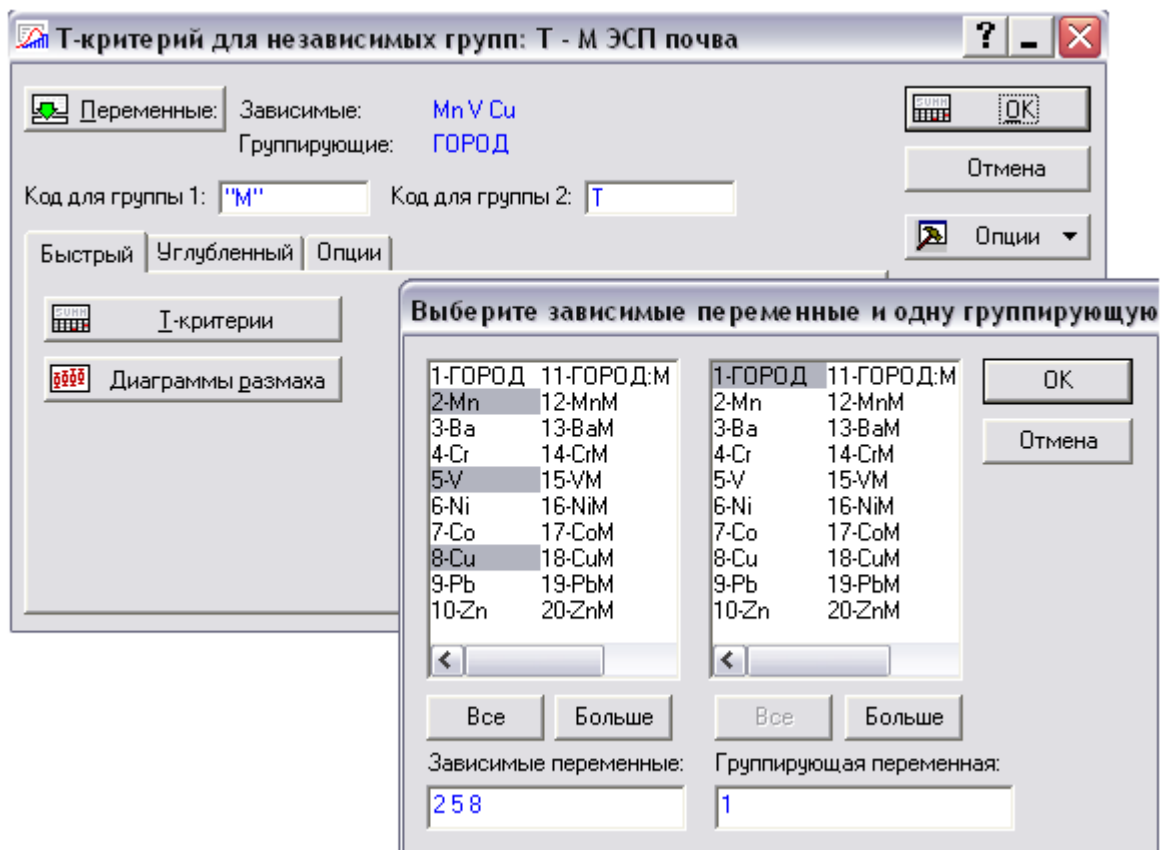


Рис.3.12. Окно  Т-критерий для независимых групп: Т - М ЭСП почва

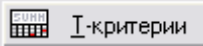
Шаг 5. После этого нажмите кнопку  I-критерии (рис.3.12). Следующая таблица с результатами появится на экране:




Таблица 3.4

Параметрические t - критерий Стьюдента (и F - критерий Фишера) сравнения содержаний микроэлементов в почво-грунтах М и Т.

	m_M	m_T	t -знач.	сс	p_t	N_M	N_T	S_M	S_T	F -знач.	p_F
Mn	0,393	0,422	-1,068	198	0,287	100	100	0,197	0,187	1,102	0,630
V	5,650	6,290	-2,102	198	0,037	100	100	2,096	2,208	1,110	0,605
Cu	5,030	5,570	-2,017	198	0,045	100	100	1,800	1,981	1,211	0,343

По табл.3.4 незначимо различаются только m_M и m_T для Mn .

Для проведения оценки различия содержаний остальных химических элементов по двум территориям посредством применения непараметрических критериев различия (неоднородности) выборок необходимо выполнить последовательность следующих действий (п. 2.2.2):

Шаг 1. Нажмите  **Непараметрическая статистика** в головном меню модуля **Анализ**, выберите  **Сравнение двух независимых групп** и нажмите  **OK** (рис. 3.13).

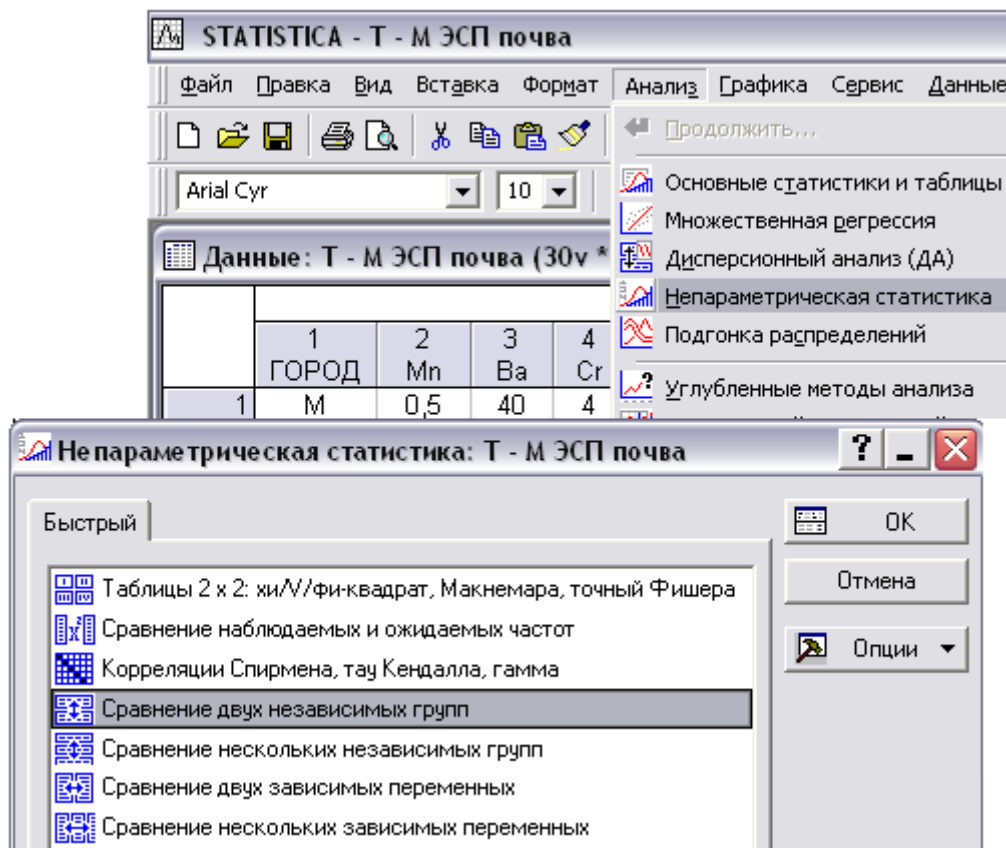






Рис.3.13. Открытие окна  **Непараметрическая статистика: Т - М ЭСП почва** в модуле **Анализ**

Шаг 2. В открывшемся окне нажмите кнопку  **Переменные** и выберите переменные категориальной базы данных (переменная «Город» - группирующая, а микроэлементы Ba, Cr, Ni, Co, Pb, Zn - зависимые) и нажмите  **OK**. После этого в окнах кодов для групп появятся «M» и «T» (рис.3.14).

Шаг 3. После этого нажмите кнопку  **Критерий Колмогорова-Смирнова** (рис.3.14).

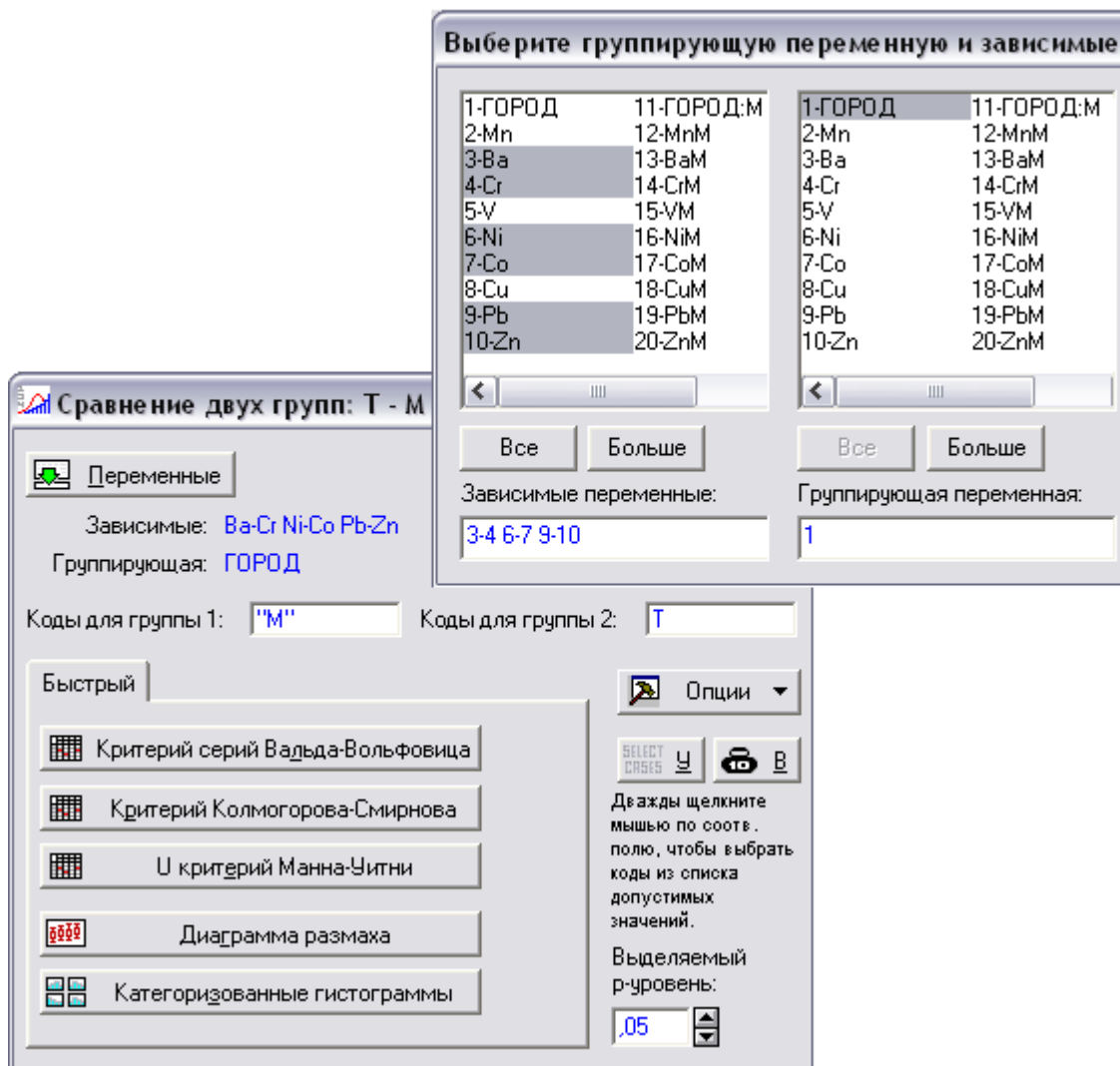



Рис.3.14. Окно  Сравнение двух групп: Т - М ЭСП почва

Следующая таблица с результатами появится на экране (табл.3.5).

Таблица 3.5

Непараметрический критерий Колмогорова-Смирнова сравнения средних содержаний микроэлементов в почво-грунтах М и Т

К-S	<i>p</i> -уровень	m_M	m_T	S_M	S_T	N_M	N_T
Ba	$p > .10$	40,80	40,40	14,953	15,565	100	100
Cr	$p < .001$	6,640	11,02	2,634	4,472	100	100
Ni	$p > .10$	3,520	3,330	1,123	1,064	100	100
Co	$p > .10$	1,713	1,695	0,927	0,861	100	100
Pb	$p > .10$	3,350	2,900	1,266	1,283	100	100
Zn	$p < .001$	9,640	6,680	3,721	3,408	100	100

Основной интерес представляет столбец уровней значимости (*p*-уровень) различий средних m_M и m_T содержаний микроэлементов в гг. М и Т. Высоко значимые различия выделены серыми заливками (**Cr** и **Zn**).

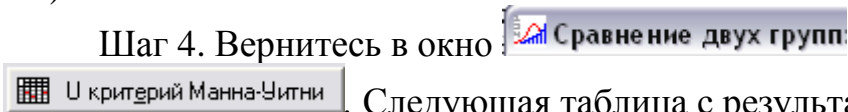
Шаг 4. Вернитесь в окно  **Сравнение двух групп** и нажмите кнопку **U критерий Манна-Уитни**. Следующая таблица с результатами появится на экране:

Таблица 3.6

Непараметрический *U*- критерий Манна - Уитни сравнения средних содержаний микроэлементов в почве М и Т

M-W	Сум. ранг ИНАА	Сум. ранг ISP	U	Z	<i>p</i> -уровень	N _П	N _С
Ba	10163	9936	4886	0,277	0,782	100	100
Cr	7159	12940	2109	-7,063	0,000	100	100
Ni	10571	9528	4478	1,274	0,203	100	100
Co	10091	10008	4958	0,101	0,919	100	100
Pb	11074	9025	3975	2,503	0,012	100	100
Zn	12669	7431	2381	6,399	0,000	100	100

Здесь также основной интерес представляет столбец уровней значимости (*p*-уровень) различий средних m_M и m_T содержаний микроэлементов в гг. М и Т. Значимые и высоко значимые различия выделены жирным шрифтом (**Cr**, **Pb** и **Zn**).

Таким образом, по совокупности результатов на основе параметрических и непараметрических критериев можно сделать вывод о том, что высоко значимыми являются различия средних содержаний химических элементов Cr (6,64 и 11,02) и Zn (9,64 и 6,68); значимыми являются различия средних содержаний химических элементов V (5,65 и 6,29), Cu (5,03 и 5,57) и Pb (3,35 и 2,90) и незначимыми являются различия средних содержаний Mn (0,393 и 0,422); Ba (40,80 и 40,40); Ni (3,52 и 3,33) и Co (1,713 и 1,695).

Полученные результаты визуально подтверждаются соответствующими составными гистограммами. Составные гистограммы позволяют выдвигать и проверять гипотезы о числовых характеристиках распределений.

Для построения составных гистограмм необходимо выполнить последовательность следующих действий (п. 2.1.3):

Шаг 1. Запустите в головном меню (полоса меню) модуль **Графика**, выберите **Категоризованные графики** в стартовой панели модуля и далее щелкните мышкой на разновидности **Гистограммы...** (рис.3.15).

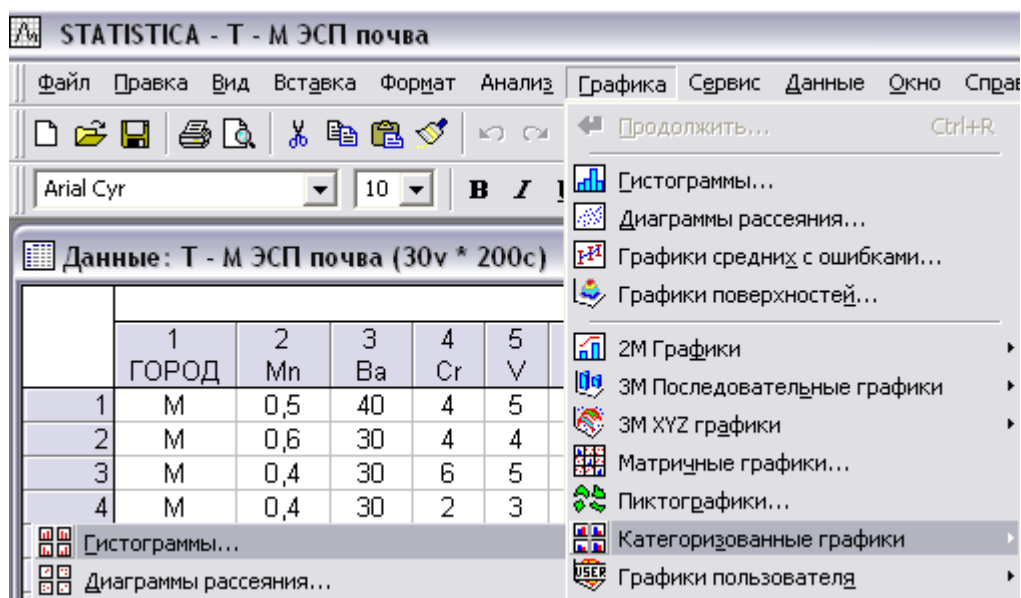


Рис.3.15. Открытие окна **2М Категоризованные гистограммы** в модуле **Графика**

Шаг 2. В появившемся окне **2М Категоризованные гистограммы** щелкните **Дополнительно** (рис.3.16).

Шаг 3. В открывшемся окне (рис.3.16) нажмите кнопку **Переменные:** и выберите переменные категориальной базы данных (например, переменная «Mn» – в окне **Переменная:** и «Город» – в окне **X-Категория:**) и нажмите ОК.

Шаг 4. Далее в окне **X-Категории** выделите режим **Коды:**, нажмите кнопку **Задать коды**, выберите **Все** в **Коды категорий:** (рис.3.16) и нажмите ОК.

Шаг 5. После этого в окне **Размещение** выделите режим **Вместе**, в окне **Параметры гистограммы** выделите режимы **Интервал между столбцами**, в окне **Интервалы** выделите режим **Авт**, в окне **Подгонка:** выберите **Нормальное** распределение (рис.3.16) и нажмите ОК.

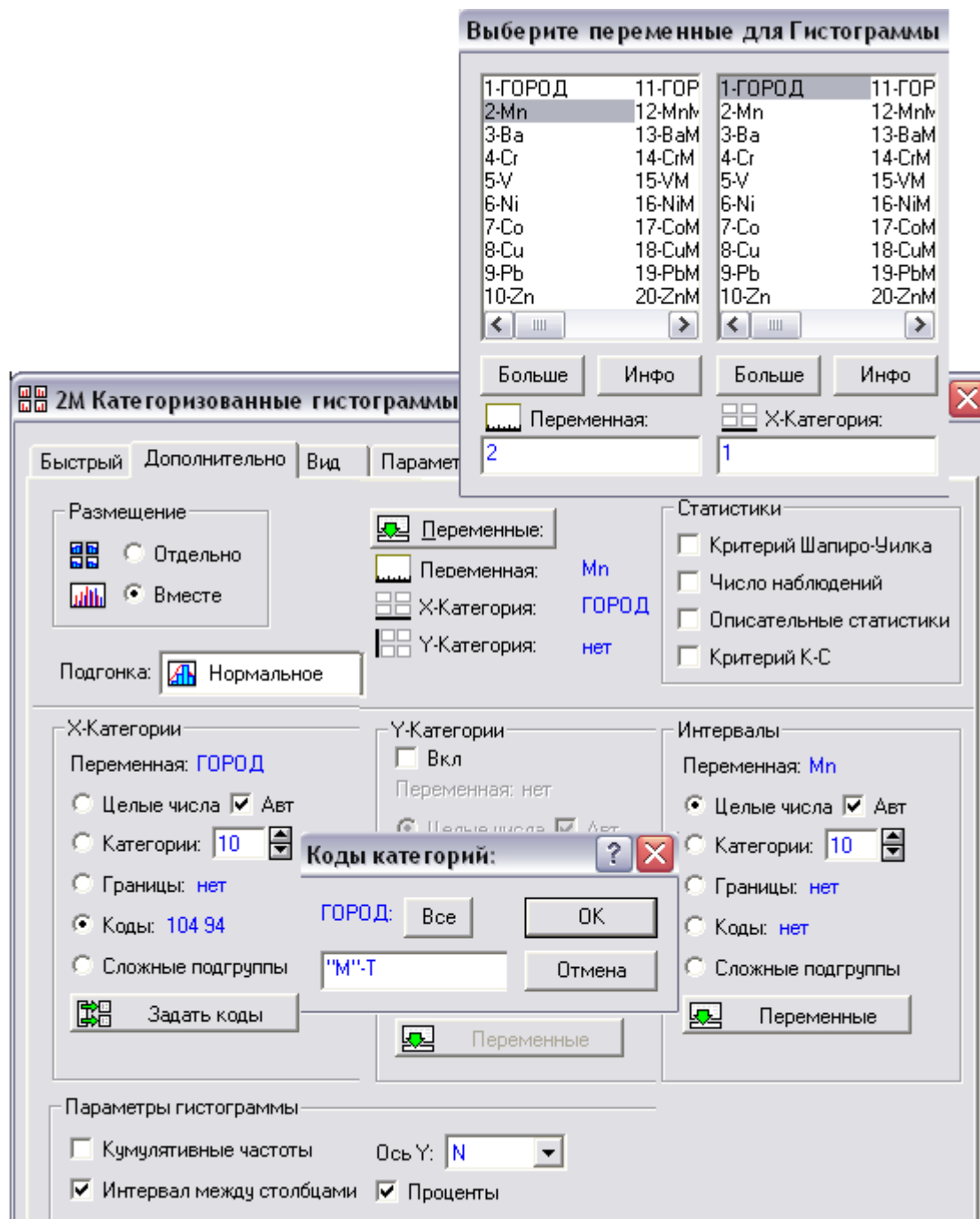


Рис. 3.16. Окно **2M Категоризованные гистограммы** для составной гистограммы Mn по городам М и Т

Следующая составная гистограмма Mn по городам М и Т (различия незначимы) появится на экране (рис.3.17). Аналогично строится составная гистограмма распределения содержания Cr (рис.3.18), Co (рис.3.19), Zn (рис.3.20) по городам М и Т.

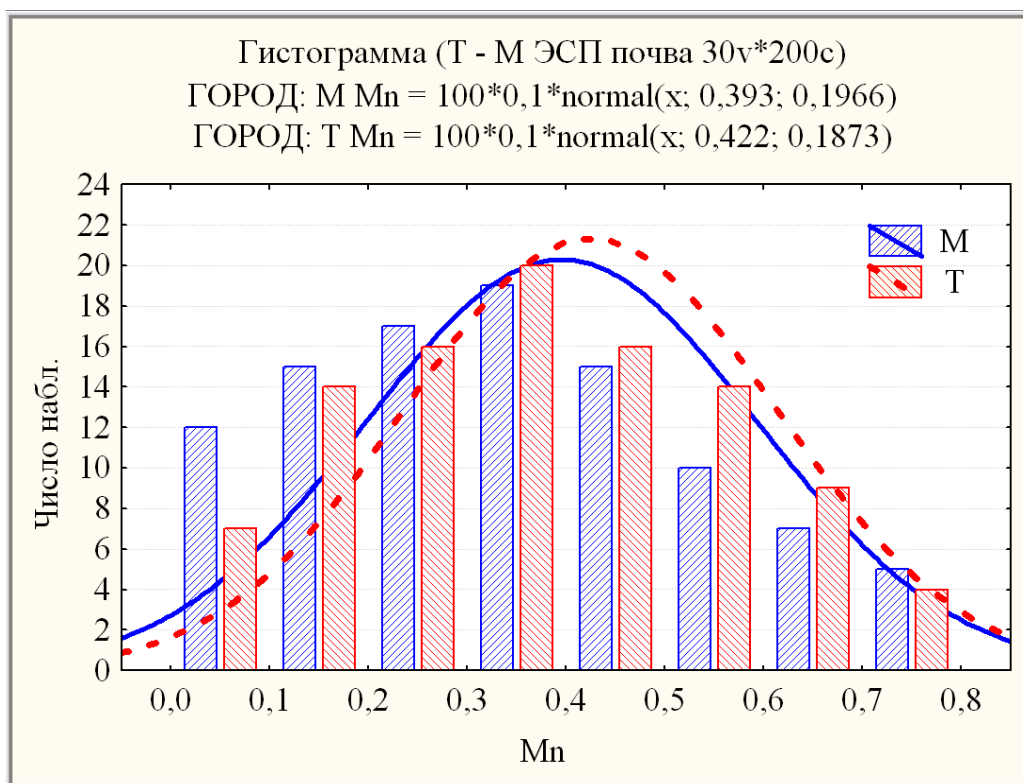


Рис. 3.17. Составная гистограмма Mn по М и Т (различия незначимы).

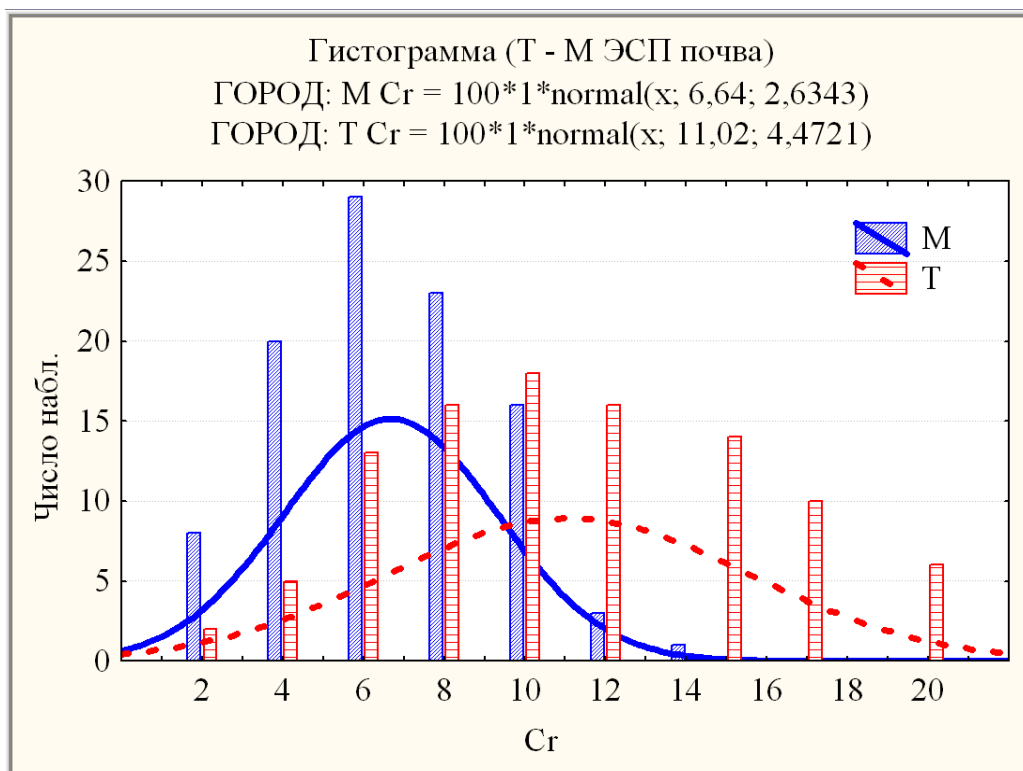


Рис. 3.18. Составная гистограмма Cr по М и Т (различия высоко значимы).

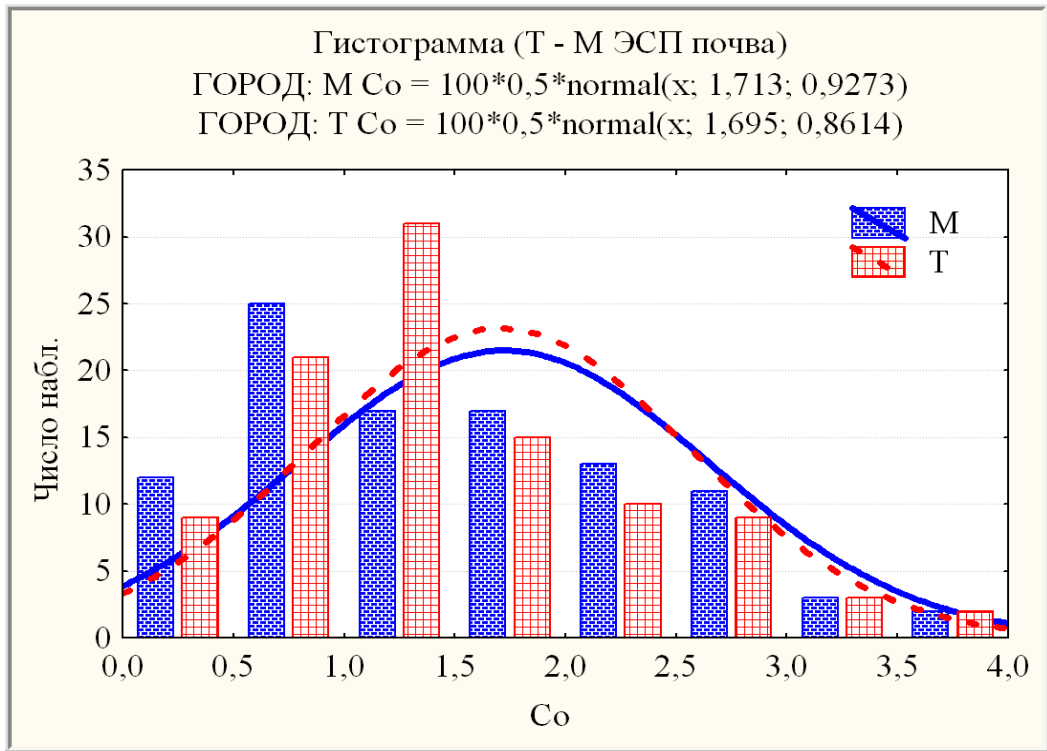


Рис. 3.19. Составная гистограмма Co по М и Т (различия незначимы).

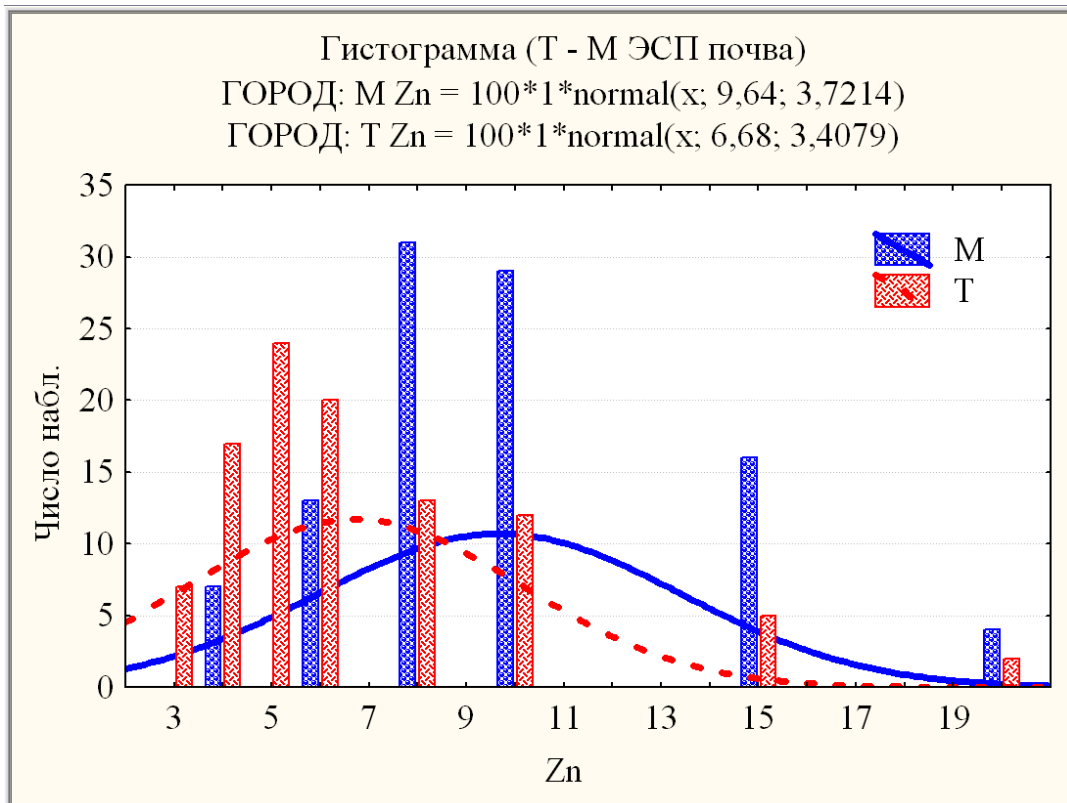


Рис. 3.20. Составная гистограмма Zn по М и Т (различия высоко значимы).

3.1.4. Корреляционный анализ содержаний микроэлементов в почве территорий.

Корреляция между результатами содержаний двух химических элементов может характеризоваться коэффициентом парной корреляции Пирсона r для нормальных распределений или ранговым коэффициентом корреляции Спирмена R , если распределения заметно отличается от нормального. При этом значимость корреляции определяется уровнем значимости отличия коэффициента корреляции от нуля.

Для вычисления корреляционной матрицы Пирсона r одной из двух ассоциаций необходимо выполнить последовательность следующих действий (п. 2.4.1):

Шаг 1. Запустите в головном меню модуль **Анализ**, затем нажмите **Основные статистики и таблицы**, выберите **Парные и частные корреляции** в появившемся окне (рис.3.21) и нажмите **OK**.

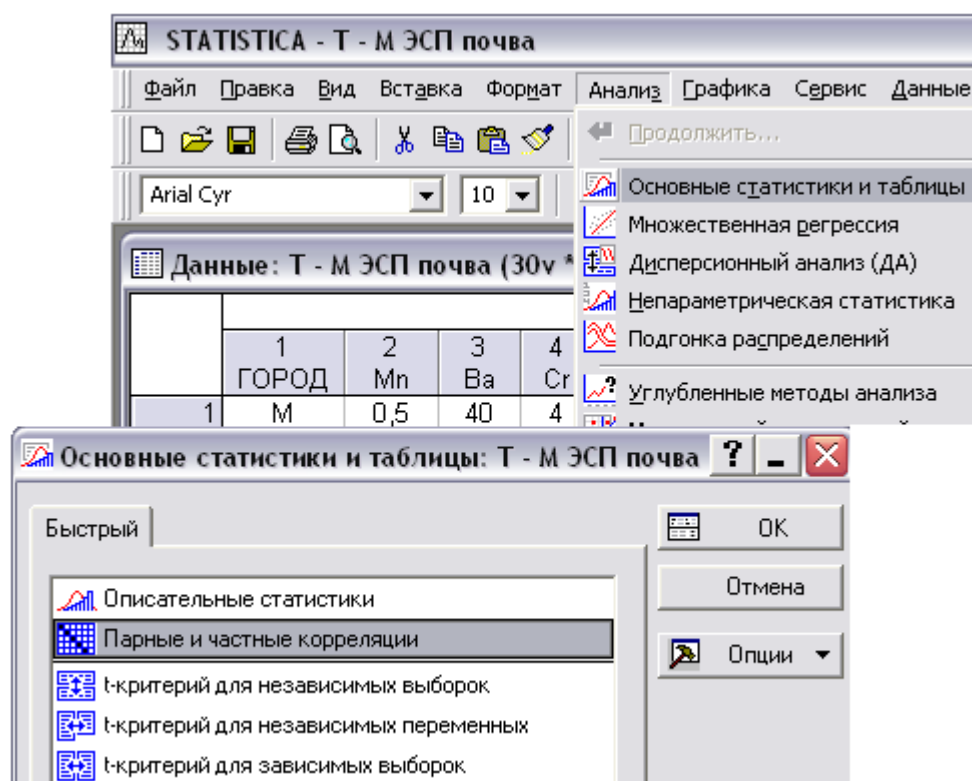


Рис. 3.21. Выбор **Парные и частные корреляции** в модуле **Анализ**

Шаг 2. В открывшемся окне **Парные и частные корреляции** нажмите левую кнопку **Квадратная матрица**, выберите все переменные М базы

данных и нажмите ОК (рис. 3.22). После этого нажмите кнопку

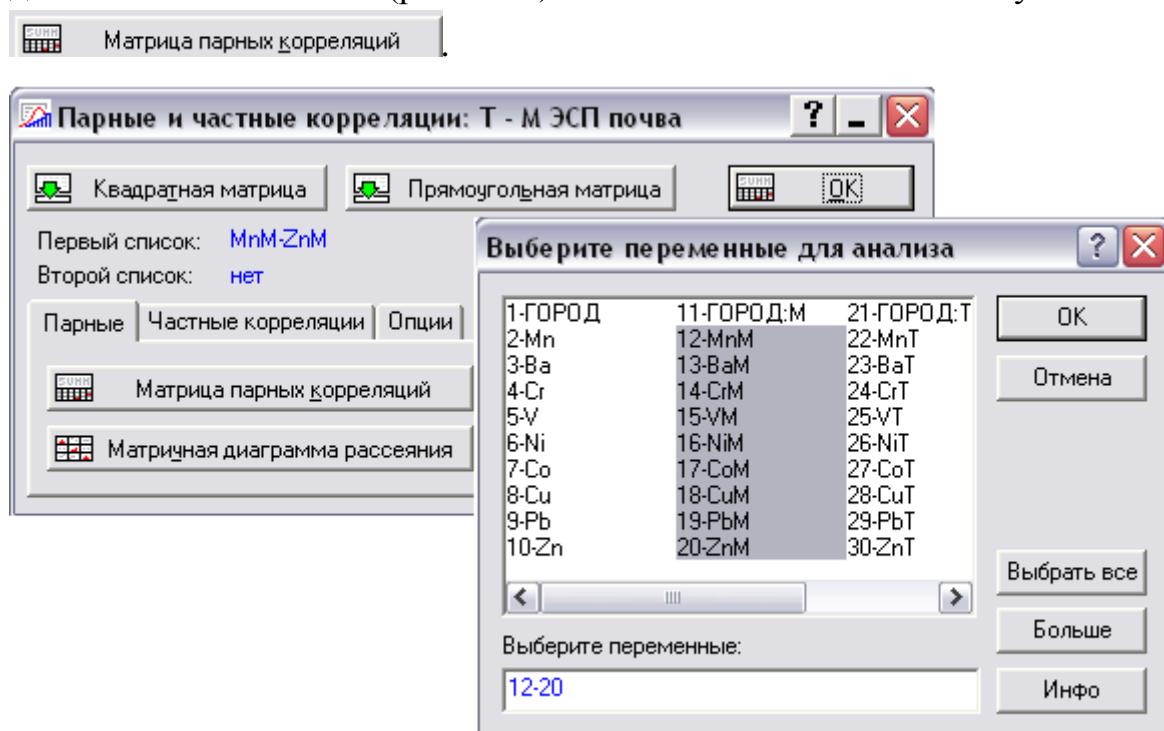


Рис. 3.22. Окно **Парные и частные корреляции**

На экране появится симметричная таблица с результатами (в табл.3.7 - правая верхняя треугольная половина, где нижняя критическая граница **0,20** значимых коэффициентов парной корреляции Пирсона и верхняя **0,68** выделены жирным шрифтом).

Таблица 3.7

Корреляционная матрица содержаний микроэлементов в почве М

	Mn _М	Ba _М	Cr _М	V _М	Ni _М	Co _М	Cu _М	Pb _М	Zn _М	
Mn _М	1,00	0,49	0,30	0,27	0,18	0,30	0,19	0,46	0,20	<i>r</i>
Ba _М	0,44	1,00	0,32	0,27	-0,02	-0,09	-0,09	0,43	-0,06	
Cr _М	0,27	0,26	1,00	0,60	0,41	0,30	0,10	0,23	0,07	
V _М	0,27	0,29	0,60	1,00	0,49	0,41	0,21	0,23	-0,01	
Ni _М	0,20	-0,01	0,45	0,46	1,00	0,44	0,10	0,24	0,13	
Co _М	0,28	-0,10	0,31	0,41	0,45	1,00	0,68	0,21	0,44	
Cu _М	0,16	-0,08	0,09	0,26	0,14	0,71	1,00	0,22	0,42	
Pb _М	0,44	0,45	0,13	0,20	0,23	0,17	0,18	1,00	0,37	
Zn _М	0,21	-0,10	0,08	-0,04	0,15	0,57	0,50	0,35	1,00	
R										

Для вычисления корреляционной матрицы Спирмена R одной из двух ассоциаций необходимо выполнить последовательность следующих действий (п. 2.4.2):

Шаг 1. Запустите в головном меню модуль **Анализ**, затем нажмите **Непараметрическая статистика**, выберите **Корреляции Спирмена**, в появившемся окне (рис.3.23) и нажмите **OK**.

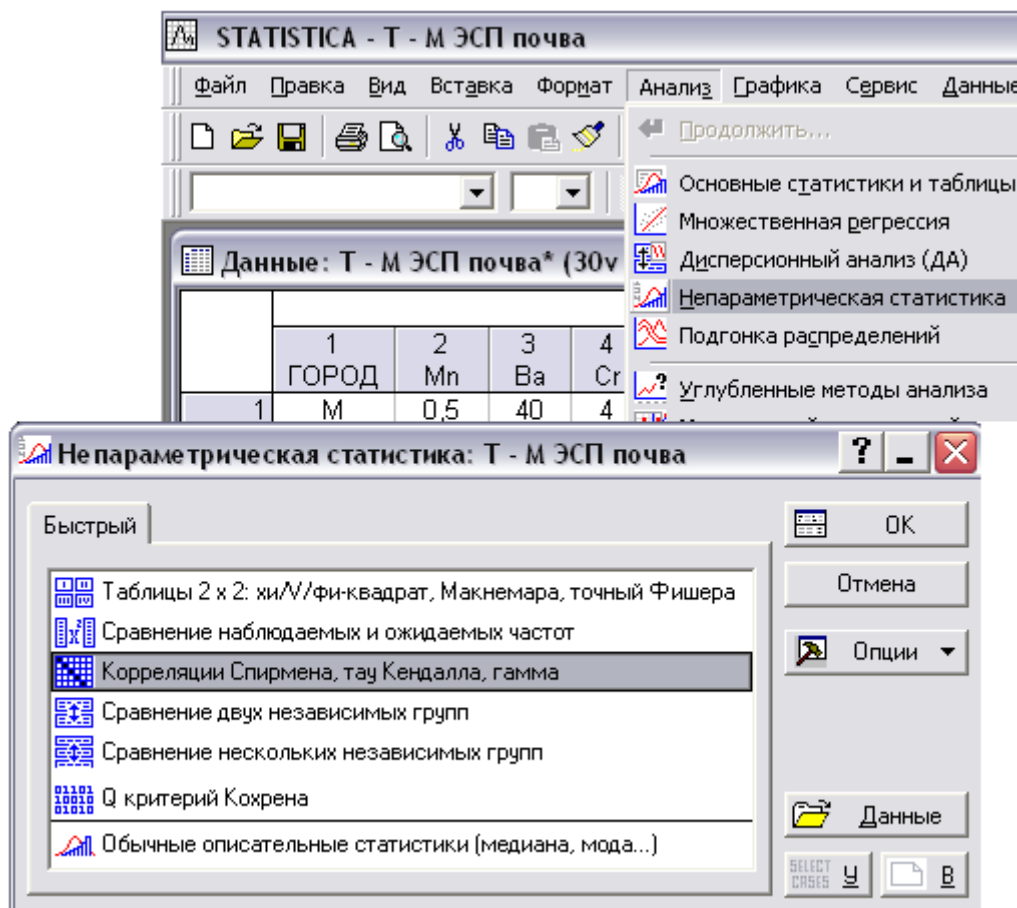


Рис. 3.23. Выбор **Корреляции Спирмена** в **Непараметрическая статистика:**

Шаг 2. В открывшемся окне **Ранговые корреляции** выберите режим вычисления **Вычислить: Квадратная матрица**, далее нажмите кнопку **Переменные** и выберите все переменные М базы данных г. М и нажмите **OK** (рис.3.24). После этого нажмите кнопку **Спирмена R**.

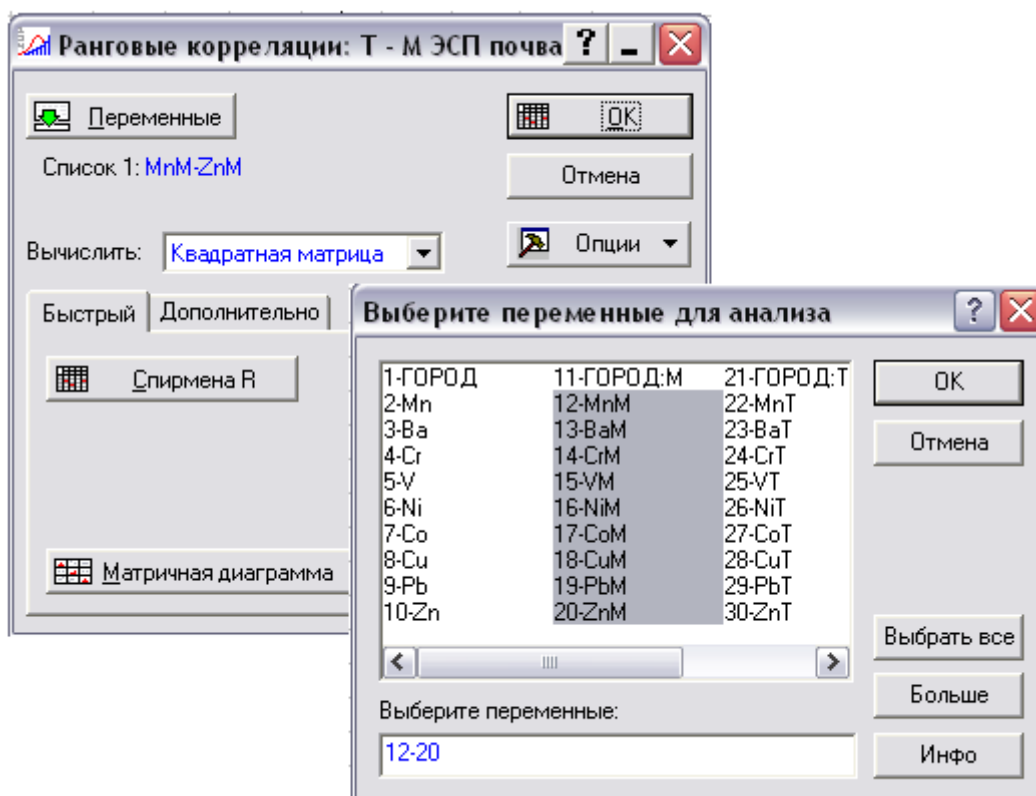


Рис. 3.24. Окно **Ранговые корреляции** в режиме **Квадратная матрица**

На экране появится симметричная таблица с результатами (в табл.3.7 - левая нижняя треугольная половина, где нижняя критическая граница **0,21** значимых коэффициентов парной корреляции Спирмена **R** и верхняя **0,71** выделены жирным шрифтом).

В данном случае наряду со значениями коэффициентов парной корреляции Спирмена **R** можно вычислить также соответствующие критерии $t(N-2)$ проверки нулевой гипотезы $H_0: \mathbf{R} = 0$ (аналогично (1.34)) и их p -уровни значимости :

Шаг 2'. В открывшемся окне **Ранговые корреляции** выберите режим вычисления **Вычислить: Подробный отчет**, далее нажмите кнопку **Переменные** и выберите два набора переменных базы данных г. М (например, переменная «Mn» – в окне **Первый список**, а все остальные переменные – в окне **Второй список**) и нажмите **OK** (рис.3.24'). После этого нажмите кнопку **Спирмена R**.

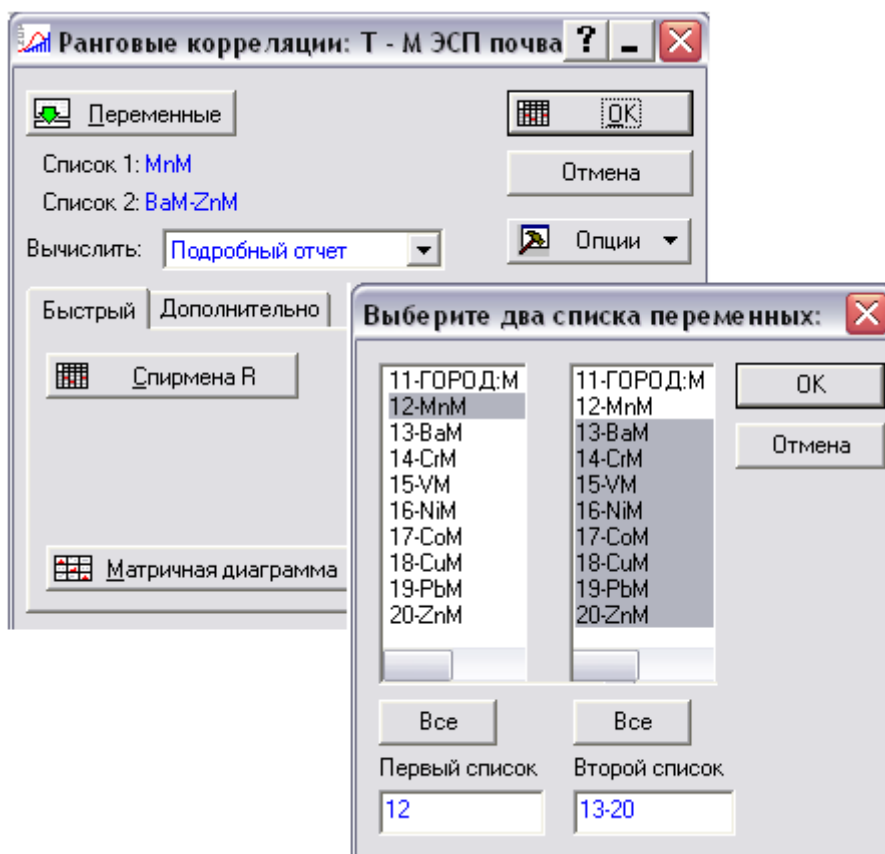


Рис. 3.24'. Окно **Ранговые корреляции** в режиме **Подробный отчет**

На экране появится первый столбец корреляционной матрицы Спирмена R в следующем формате (табл.3.8):

Таблица 3.8
Первый столбец корреляционной матрицы Спирмена выборки содержа-
ний микроэлементов в почво-грунтах М

	N	Спирмена R	$t(N-2)$	p -уровень
Mn_M & Ba_M	100	0,440675	4,859770	0,000004
Mn_M & Cr_M	100	0,268415	2,758402	0,006931
Mn_M & V_M	100	0,267574	2,749089	0,007116
Mn_M & Ni_M	100	0,199177	2,012068	0,046957
Mn_M & Co_M	100	0,280709	2,895286	0,004671
Mn_M & Cu_M	100	0,159763	1,602149	0,112341
Mn_M & Pb_M	100	0,444506	4,912378	0,000004
Mn_M & Zn_M	100	0,211460	2,141776	0,034690

Аналогичным образом можно заполнить остальные столбцы левой нижней треугольной половины табл.3.7.

Аналогично строится корреляционная матрица ассоциации микроэлементов в почво-грунтах территории г. Томска (табл.3.9).


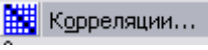
Таблица 3.9

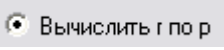
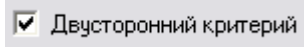
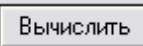
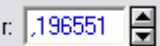
Корреляционная матрица ассоциации микроэлементов в почво-грунтах территории г. Томска (Т).

	Mn _T	Ba _T	Cr _T	V _T	Ni _T	Co _T	Cu _T	Pb _T	Zn _T	
Mn _T	1,00	0,33	0,22	0,36	0,31	0,16	-0,06	0,07	0,16	<i>r</i>
Ba _T	0,30	1,00	0,13	0,21	0,12	0,04	0,02	0,13	0,14	
Cr _T	0,25	0,18	1,00	0,45	0,37	0,21	-0,17	-0,09	-0,22	
V _T	0,33	0,22	0,48	1,00	0,47	0,59	-0,04	0,10	-0,04	
Ni _T	0,36	0,15	0,36	0,55	1,00	0,47	0,10	0,25	0,14	
Co _T	0,21	0,10	0,22	0,62	0,44	1,00	0,38	0,21	0,00	
Cu _T	-0,03	0,13	-0,15	-0,03	0,07	0,39	1,00	0,47	0,24	
Pb _T	0,10	0,18	-0,04	0,13	0,20	0,21	0,45	1,00	0,44	
Zn _T	0,20	0,21	-0,16	0,02	0,23	0,00	0,25	0,41	1,00	
R										

Жирным шрифтом в табл.3.9 выделен наименьший из значимых коэффициентов парной корреляции Пирсона $r = 0,21$ ($R = 0,20$). В табл. 3.9 выделен также жирным шрифтом наиболее значимый коэффициент корреляции $r = 0,59$ ($R = 0,62$).

В системе STATISTICA предоставлена возможность вычисления критического коэффициента корреляции Пирсона по заданному уровню значимости и объему выборки. Для этого необходимо:

Шаг 1. Запустите в головном меню модуль **Анализ**, нажмите клавишу  в стартовой панели модуля и в появившемся окне выберите  (рис.3.25)

Шаг 2. В окне **Значимость коэффициента корреляции Пирсона** введите значения объема выборки $N = 100$ и уровень значимости $p = 0,05$, выделите точкой  и галочкой  и нажмите кнопку  (рис. 3.26). В окне г:  появится значение $r =$

0,196551 - критическое значение – граница значимых, то есть отличных от нуля на уровне значимости 0,05 при объеме выборки $N = 100$, коэффициентов парной корреляции Пирсона.

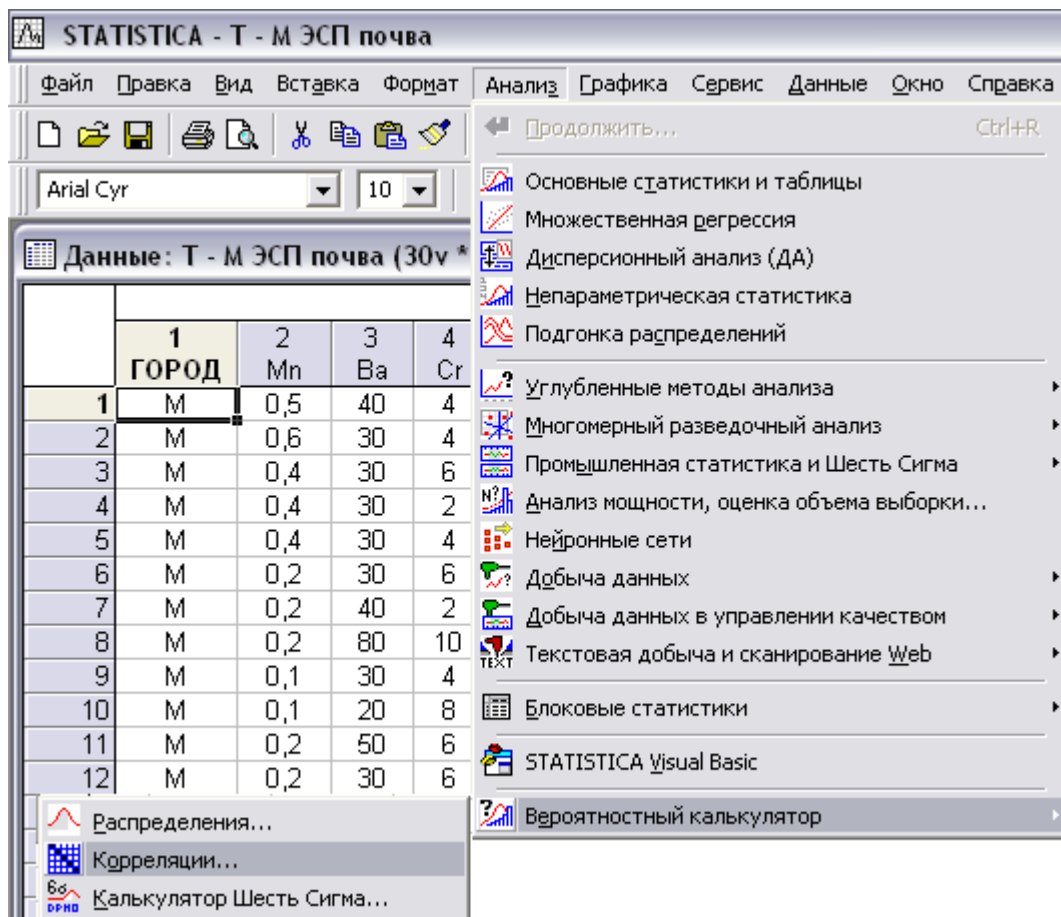


Рис. 3.25. Выбор  Корреляции... в  Вероятностный калькулятор модуля  Анализ

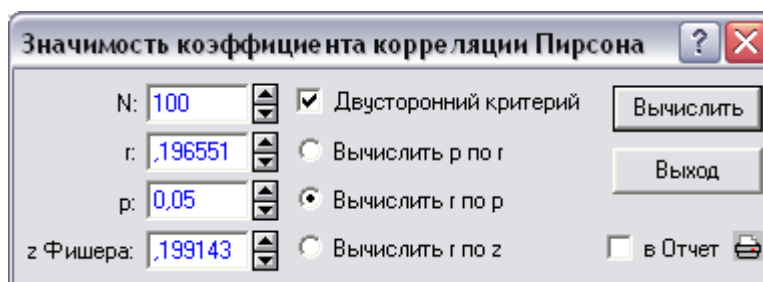


Рис. 3.26. Вычисление коэффициента корреляции по объему выборки при заданном уровне значимости.

Представляет интерес оценка значимости различия между двумя значимыми ($> 0,196551$) коэффициентами корреляции одной и той же

пары микроэлементов в разных городах (например, Mn и V в разных городах). Для этого необходимо:

Шаг 1. Запустите в головном меню модуль **Анализ**, выберите **Основные статистики и таблицы**, затем выберите **Другие критерии значимости** (рис. 3.27) и нажмите **ОК**.

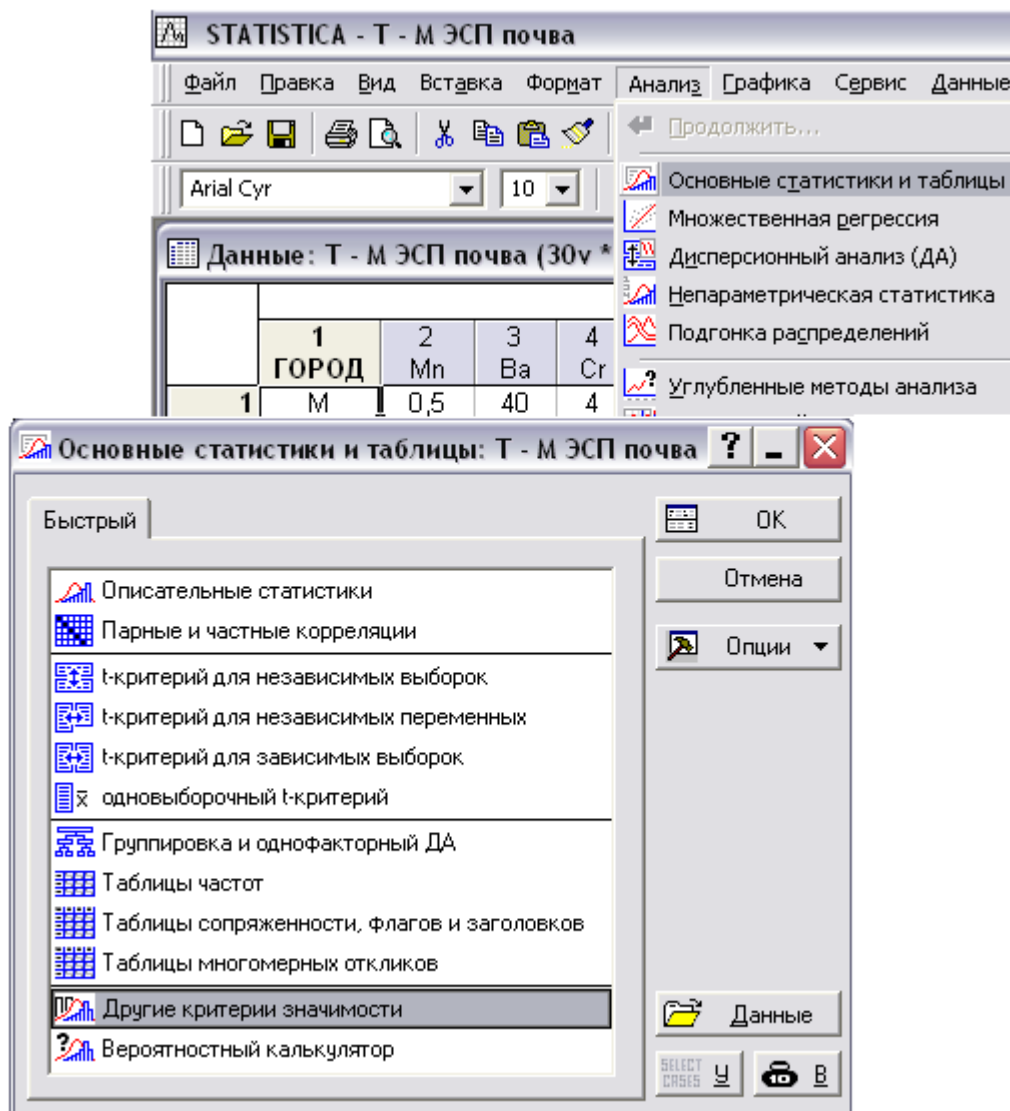


Рис. 3.27. Выбор **Другие критерии значимости** в **Основные статистики и таблицы** модуля **Анализ**

Шаг 2. В окне **Другие критерии значимости** введите значения сравниваемых значений коэффициентов корреляции (напр., $r_1 = 0,27$ и $r_2 = 0,36$), объемов выборок (напр., $N_1 = 100$ и $N_2 = 100$), выделите точкой

двухсторонний режим и нажмите **Вычислить** (рис. 3.28). Здесь же в окне появится значение $p = 0,4869$, то есть данные $r_1 = 0,27$ и $r_2 = 0,36$ при $N_1 = 100$ и $N_2 = 100$ различаются незначимо.

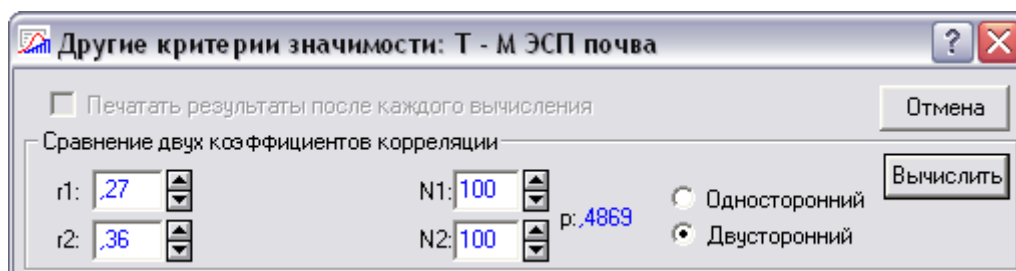


Рис. 3.28. Оценка расхождения между двумя коэффициентами корреляции Пирсона

Перебирая таким образом все значимые пары (см. табл. 3.7 и 3.9), построим таблицу значимости различия между двумя значимыми (**> 0,196551**) коэффициентами корреляции одной и той же пары микроэлементов в разных городах (табл.3.10).

Таблица 3.10

Оценка значимости различия между двумя значимыми коэффициентами корреляции одной и той же пары микроэлементов в городах М и Т.

	Mn	Ba	Cr	V	Ni	Co	Cu	Pb	Zn	
Mn				0,487						<i>r</i>
Ba	0,259									
Cr	0,881									
V	0,647	0,603	0,237							
Ni	0,227		0,454	0,400						
Co	0,604		0,501	0,045	0,172					
Cu						0,001				
Pb					0,827					
Zn	0,942						0,042	0,626		
	R									

На примере полных табл. 3.7 и 3.9 можно проследить различие между значениями коэффициентов корреляции r и R одной и той же пары микроэлементов в городах М и Т. При этом следует иметь в виду, что коэффициент корреляции Пирсона r отражает степень линейной связи для нормальных распределений, а коэффициент корреляции Спирмена R для распределений, заметно отличающихся от нормально-

го. Таким образом, напр., в табл. 3.10 в правой верхней треугольной половине осталось с учетом результатов табл.3.3 всего одно корректное значение $p = 0,487$ для нормально распределенной пары Mn & V.

Как следует из табл.3.10, из всех значимых корреляционных связей только три пары V & Co , Co & Cu и Cu & Zn значимо различаются по городам.

Согласно табл. 3.7 и 3.9, наиболее значимые корреляционные связи в выборках наблюдаются в почво-грунтах территории г. Междуреченска между Co и Cu, а в почво-грунтах территории г. Томска между Co и V. Рассмотрим эти наиболее значимые корреляционные связи.

3.1.5. Регрессионный анализ содержаний микроэлементов в почве территорий.

При построении составных диаграмм рассеяния и уравнений регрессий для наиболее значимых корреляционных связей в ассоциациях необходимо выполнить последовательность следующих действий (п. 2.4.3):

Шаг 1. Запустите в головном меню модуль **Графика**, затем выберите **Категоризованные графики** и далее щелкните мышкой на разновидности **Диаграммы рассеяния...** (рис. 3.29).

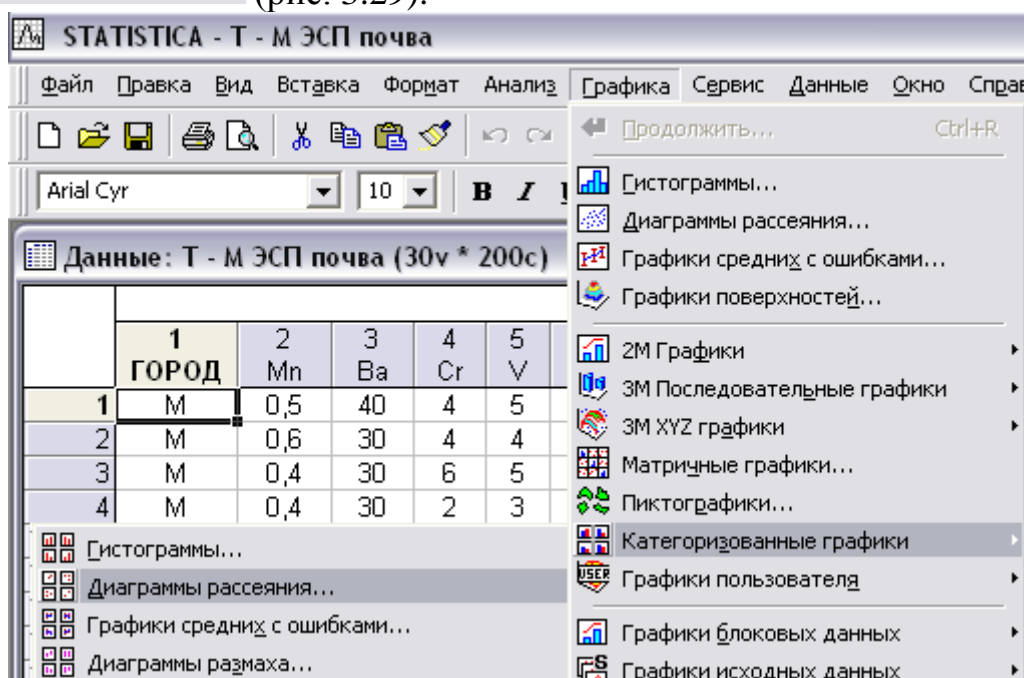


Рис. 3.29. Выбор **Диаграммы рассеяния...** в модуле **Графика**.

Шаг 2. В появившемся окне щелкните **Дополнительно** (рис. 3.30). В открывшемся окне нажмите кнопку **Переменные:** и выберите переменные категориальной базы данных (например, переменная «Со» – в окне **X:**, переменная «Си» – в окне **Y:** и переменная «Город» – в окне **X-Категория:**) и нажмите **OK**. Далее в окне **X-Категории** выделите режим **Коды:**, нажмите кнопку **Задать коды**, выберите **Все** в окне **Коды категорий:** и нажмите **OK**. После этого в окне **Размещение** выделите режим **Вместе**, в окне **Статистики** – режим **Корреляция и р-уров.**, в окне **Тип подгонки:** – режим **Линейная** (рис. 3.30) и нажмите **OK**.

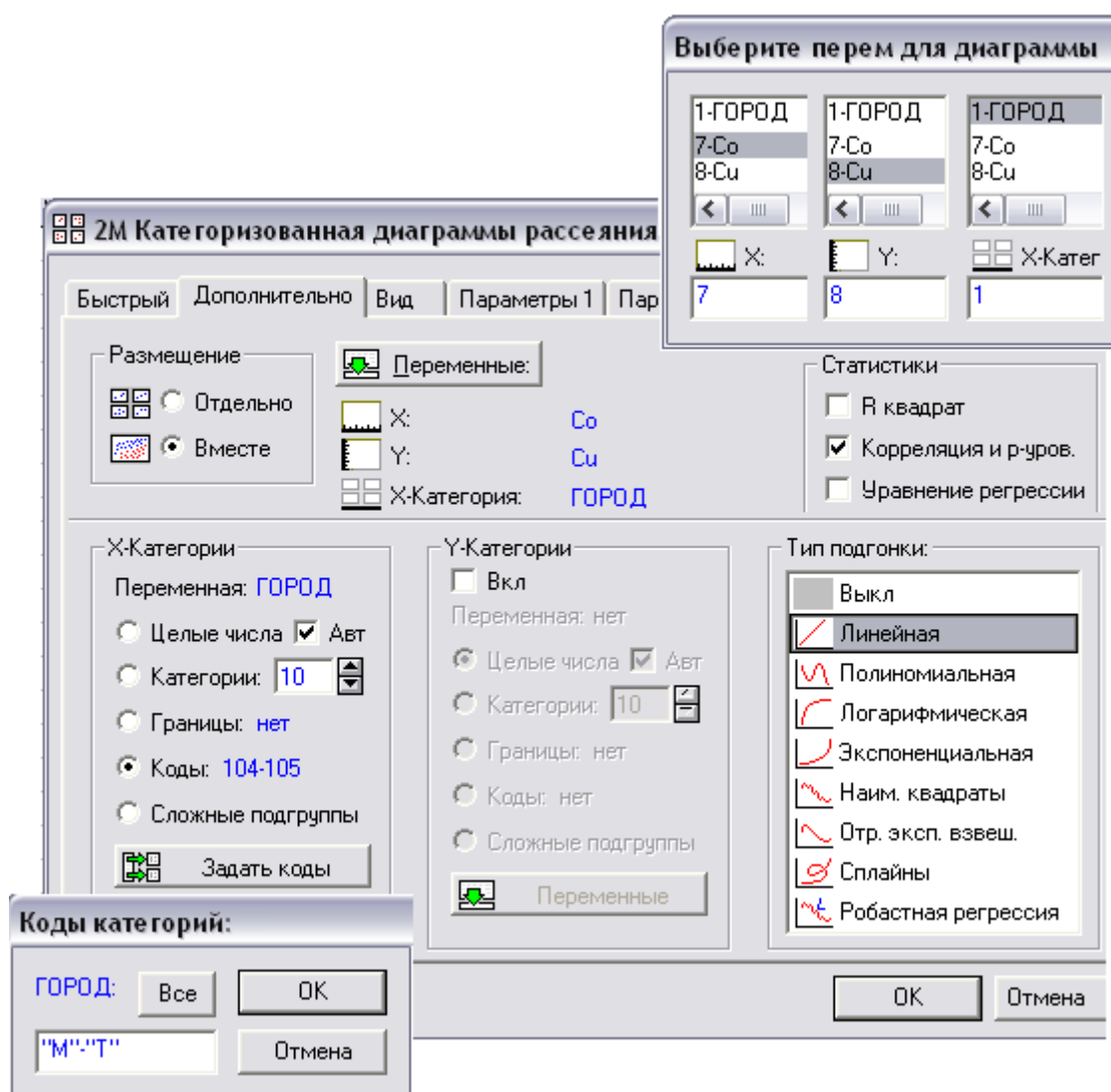


Рис. 3.30. Окно **2M Категоризованная диаграммы рассеяния**

Следующая составная диаграмма рассеяния с прямыми линиями составных регрессий $Co : Cu$ по городам М и Т появится на экране (рис.3.31). Попутно приводятся коэффициенты корреляции Пирсона с указанием их уровней значимости и уравнения регрессии зависимости Co и Cu по городам М и Т:

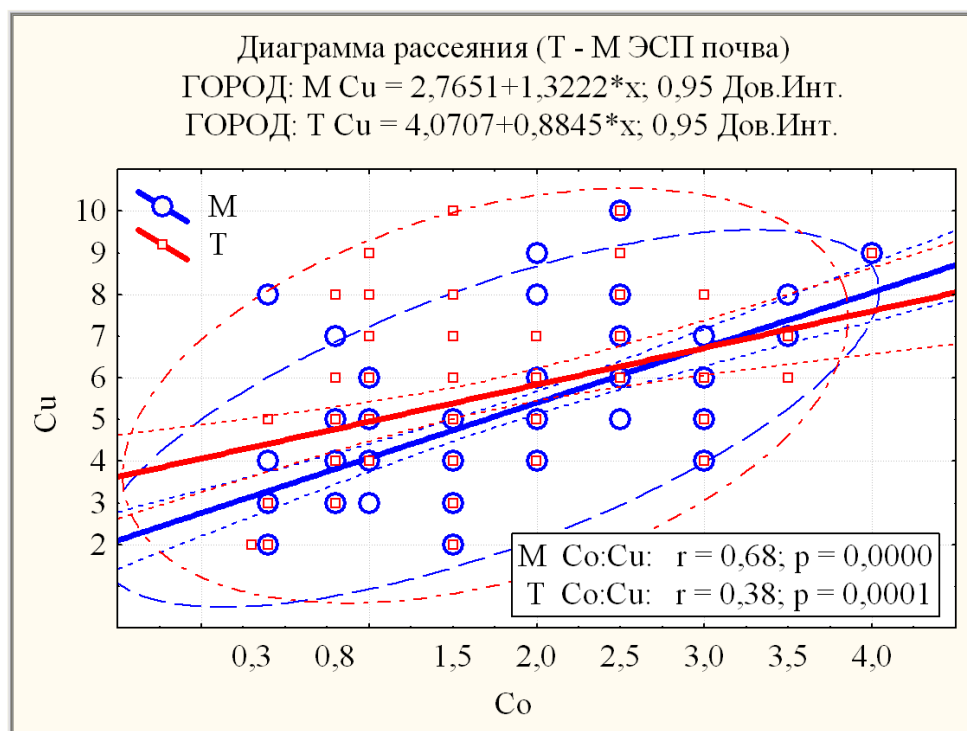


Рис.3.31. Составная диаграмма рассеяния зависимости Co и Cu по городам М и Т

В данном случае к прямой регрессии можно добавить 95% эллипс рассеяния и границы 95% доверительного интервала. Для этого достаточно щелкнуть дважды мышкой в поле диаграммы, в появившемся окне

нажать кнопку **Все параметры графика** нажать кнопку **График: Эллипс**, затем нажать кнопку **Новый эллипс**; аналогично нажать кнопку **График: Регрес. границы**, затем нажать кнопку **Новая пара границ** и нажать **ОК**.

Визуально, согласно диаграмме рассеяния, незначимые корреляционные связи между парой микроэлементов характеризуются облаком значений пары, достаточно сильно распыленным около прямой регрессии, и широкой полосой регрессии. С увеличением значения коэффициента корреляции увеличивается угол наклона прямой регрессии и усиливается группировка точек вокруг этой прямой.

Аналогичным образом строится составная диаграмма рассеяния между Co и V по городам M и T : (рис.3.32).

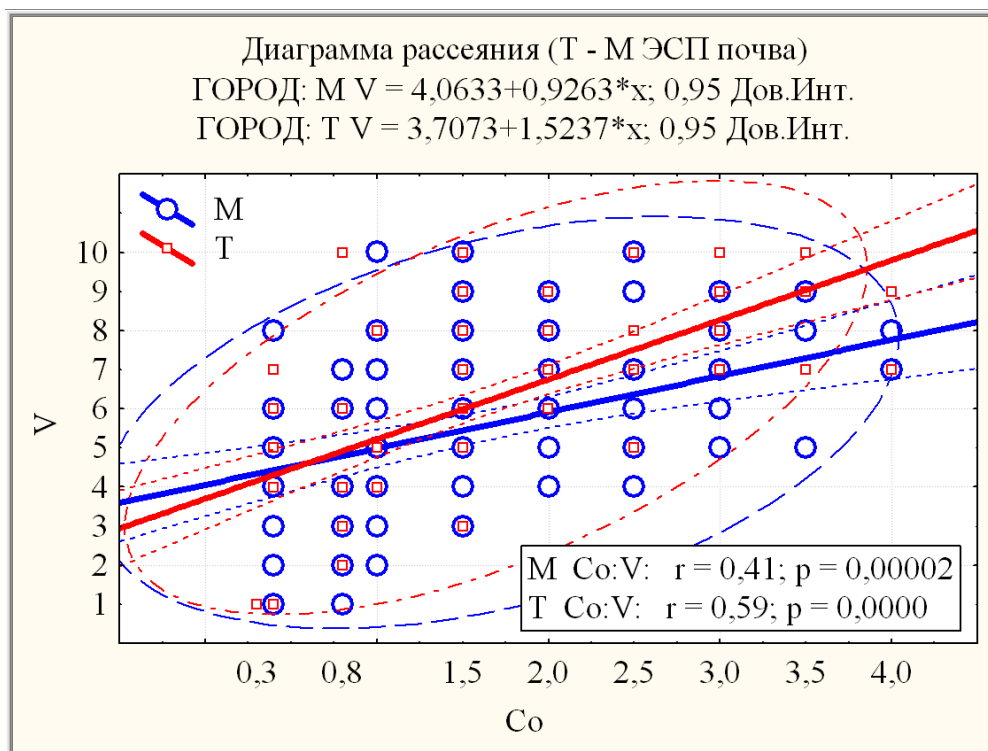
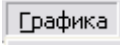
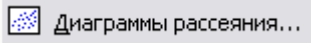
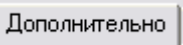




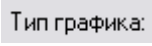
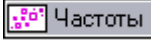
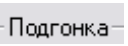
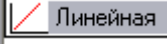
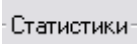


Рис.3.32. Составная диаграмма рассеяния зависимости Co и V по городам M и T

Особенностью рассматриваемых выборок является наличие одинаковых по результатам проб. Для учета частотности проб можно строить частотные диаграммы рассеяния распределения содержаний пары микроэлементов по городам. В связи с этим необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль  и щелкните мышкой на  (рис. 3.29). В появившемся окне щелкните .

Шаг 2. В открывшемся окне нажмите кнопку  и выберите переменные базы данных M , например, переменная « Co_M » – в левом окне , переменная « V_M » – в правом окне  (рис. 3.33) и нажмите .

Шаг 3. Далее в окне  выделите режим , в окне  выделите режим , в окне  выделите режим

Корреляция и r-уровень, закажите эллипс и полосы регрессии в соответствующих окнах (рис. 3.33) и нажмите .

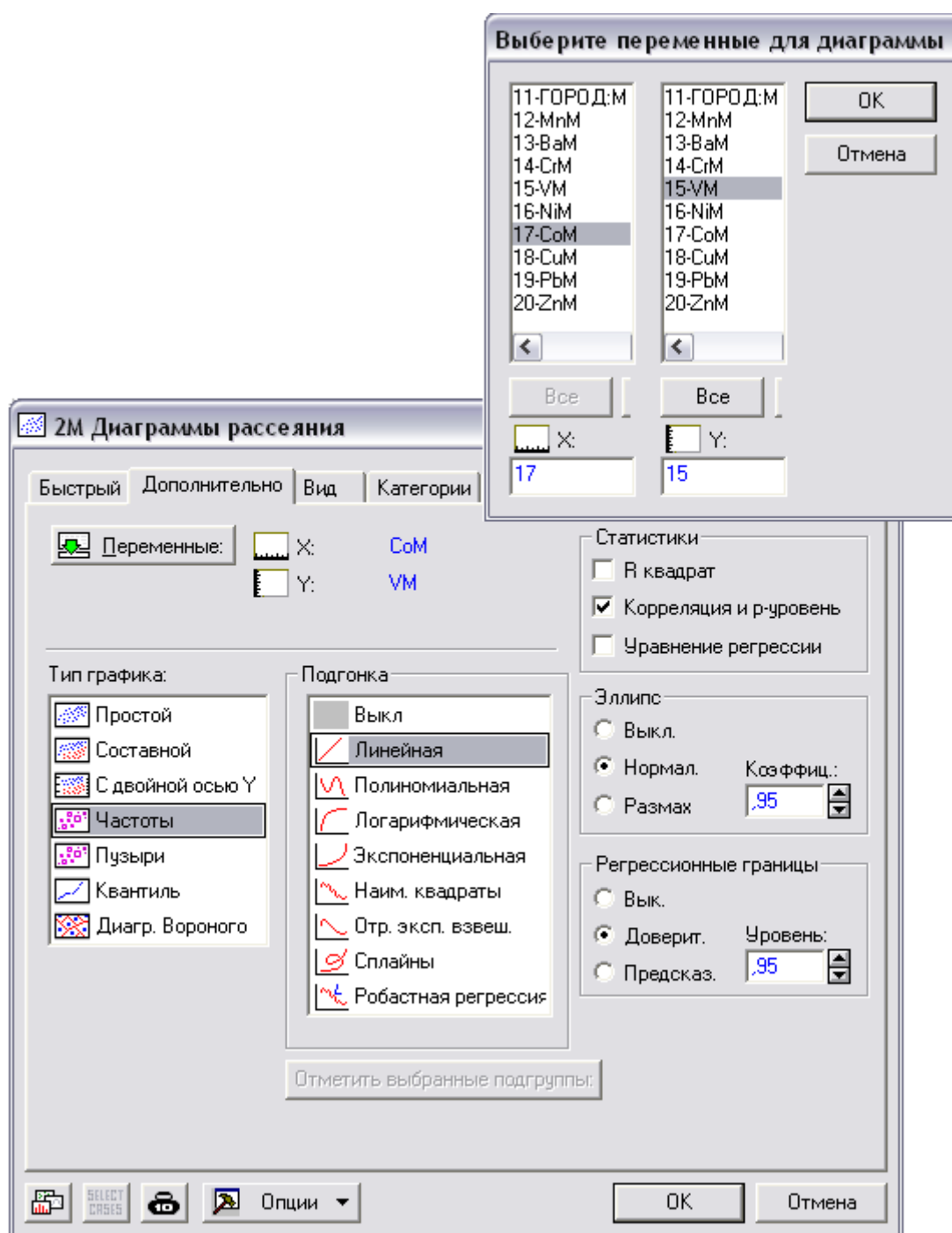


Рис. 3.33. Окно

Следующая частотная диаграмма рассеяния появится на экране (рис.3.34).

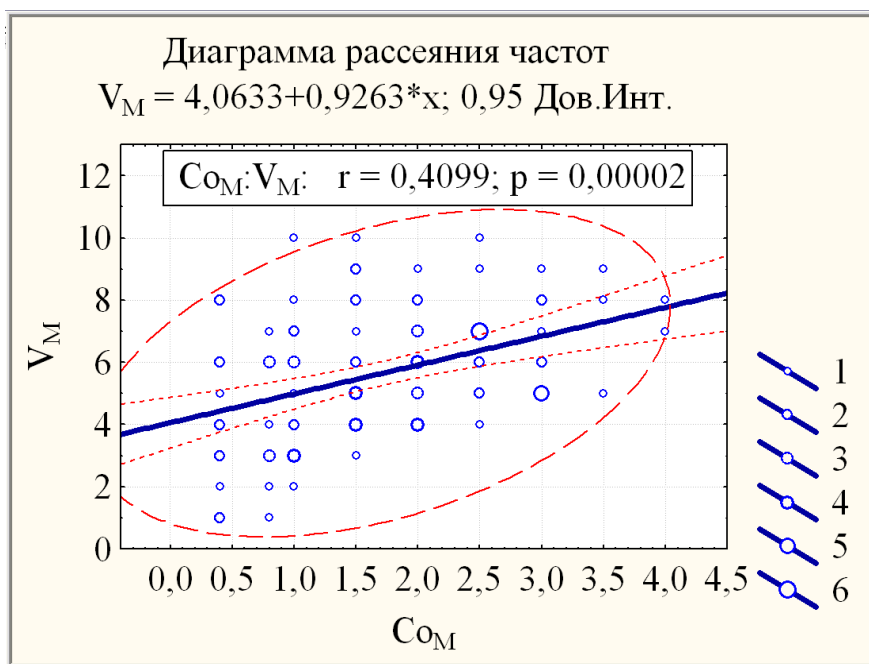


Рис. 3.34. Частотная диаграмма рассеяния $Co_M : V_M$

Аналогичным образом строится частотная диаграмма рассеяния между Co_T и V_T (рис.3.35).

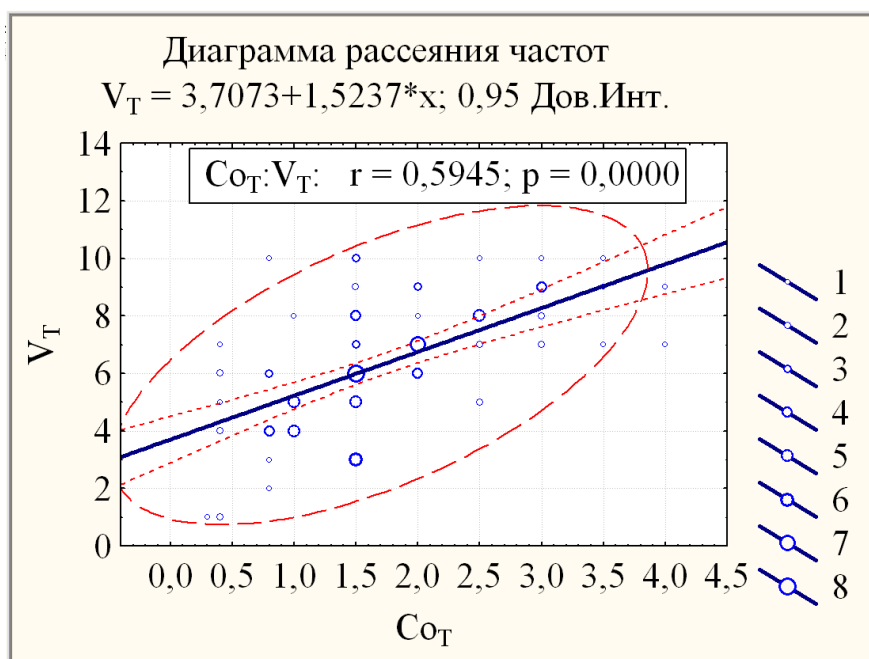







Рис. 3.35. Частотная диаграмма рассеяния $Co_T : V_T$

При желании можно отредактировать рисунки (изменить шрифт, маркировку и т. д.), щелкнув дважды мышкой по рисунку для открытия панели «Все параметры».

3.1.6. Кластерный анализ содержаний микроэлементов в почве территорий.

Для проведения кластерного анализа ассоциации микроэлементов (п.2.6) выполним последовательность следующих действий:

Шаг 1. Выберите  Многомерный разведочный анализ в стартовой панели модуля  Анализ и далее щелкните мышкой на  Кластерный анализ (рис.3.36). В появившемся окне выделите разновидность  Иерархическая классификация и щелкните  (рис. 3.36).

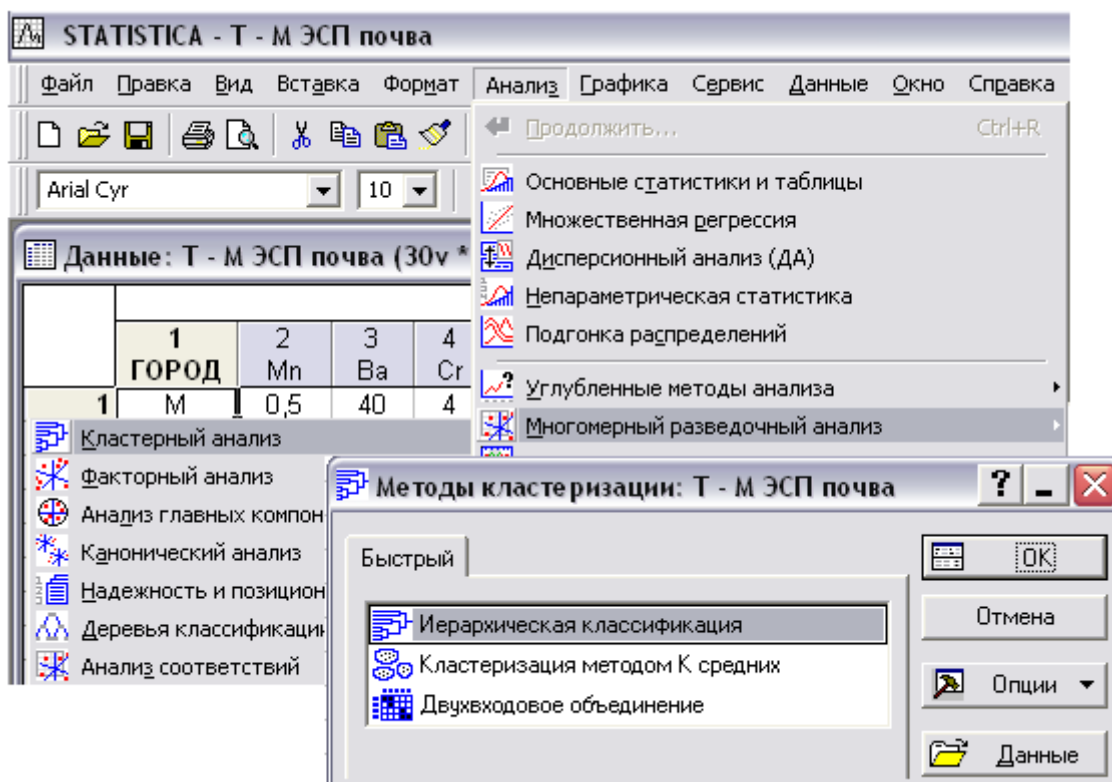












Рис. 3.36. Выбор  Методы кластеризации: в модуле  Анализ

Шаг 2. Нажмите кнопку  Переменные: в открывшемся окне  Кластерный анализ: иерархическая классификация: и выберите переменные базы данных M (рис. 3.37) и нажмите .

Шаг 3. Далее в окне  Объекты: выделите  Переменные (столбцы), в окне  Мера близости: —  1-г Пирсона, а в окне  Правило объединения: —  Метод полной связи и нажмите  (рис. 3.37).

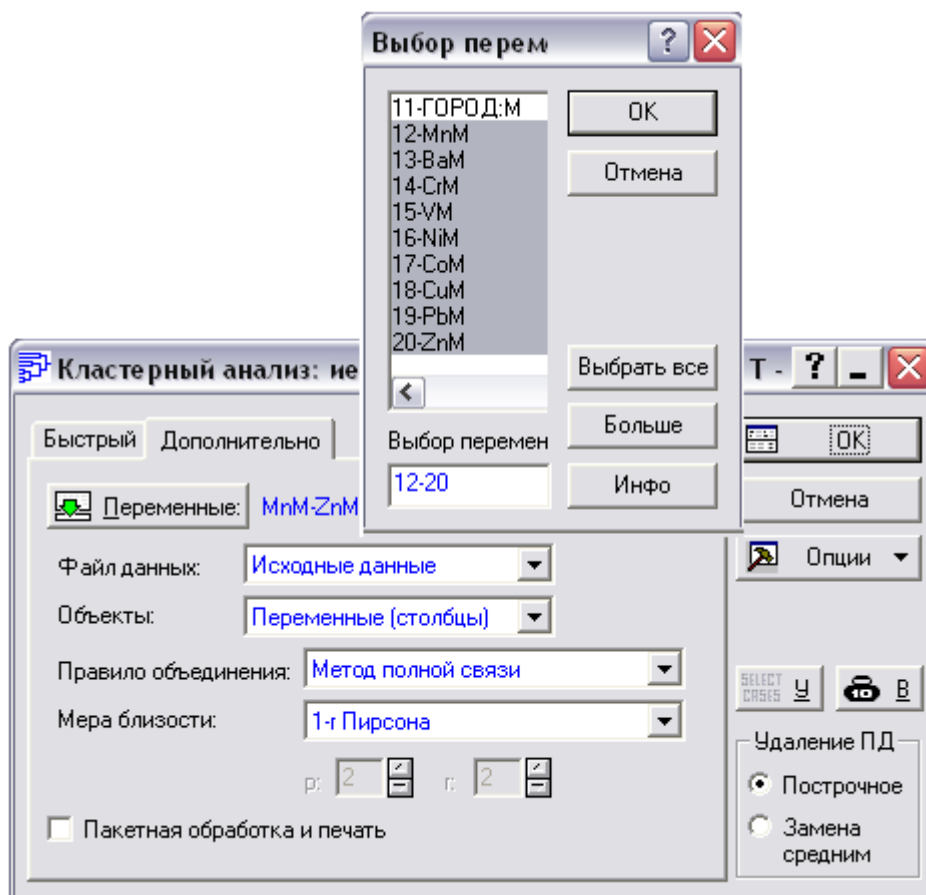


Рис. 3.37. Окно **Кластерный анализ: не** рархическая классификация:

Шаг 4. После этого щелкните **Вертикальная дендрограмма** в открывшемся окне **Результаты не рархической классификации:** рис. 3.38).

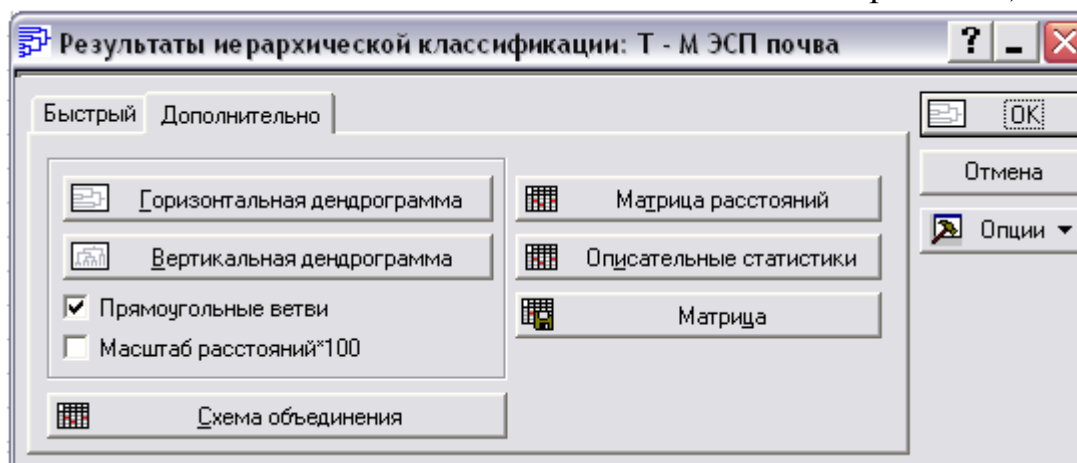


Рис. 3.38. Окно **Результаты не рархической классификации:**
Следующая дендрограмма по г. М появится на экране (рис.3.39).

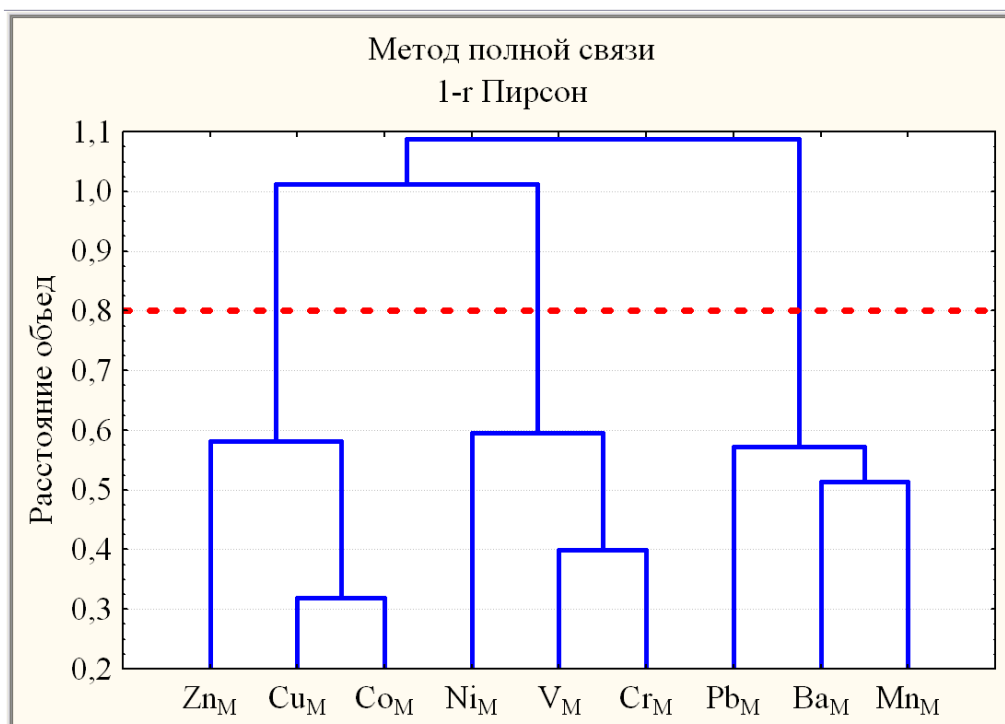


Рис. 3.39. Дендрограмма корреляционной матрицы ассоциации микроэлементов в почве г. Междуреченска ($1 - r_{0,05} \approx 0,80$; $n = 100$)

При проведении кластеризации микроэлементов в качестве меры близости микроэлементов использовано корреляционное расстояние, так как корреляционный анализ данной ассоциации микроэлементов показал значимые корреляционные связи. При этом допустимы разные методы объединения кластеров. Графические результаты кластерного анализа микроэлементов М (рис.3.39 - 41) демонстрируют устойчивое разбиение 9-ти микроэлементов на 3 значимых кластера.

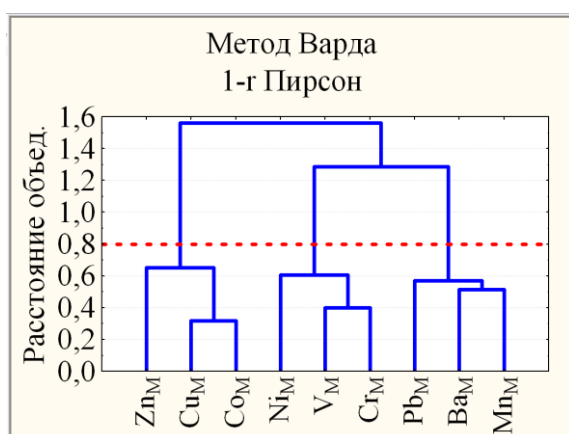


Рис. 3.40. Дендрограмма для М.
Метод Варда

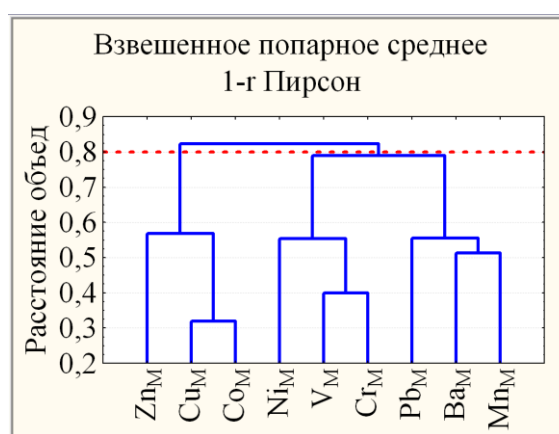
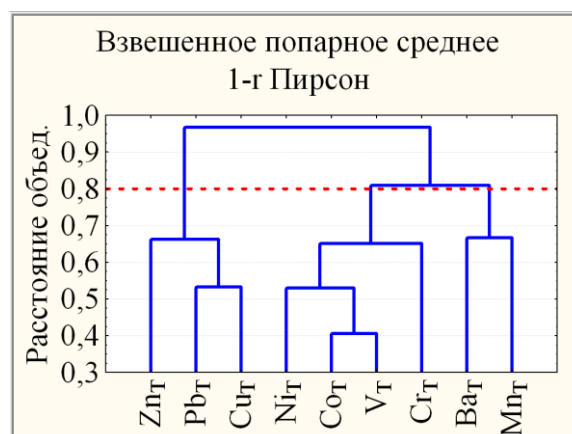
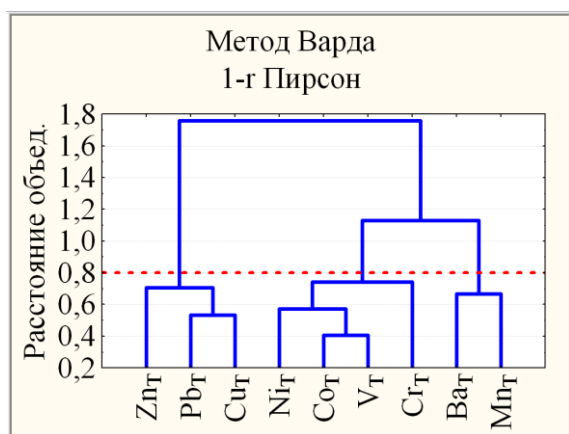
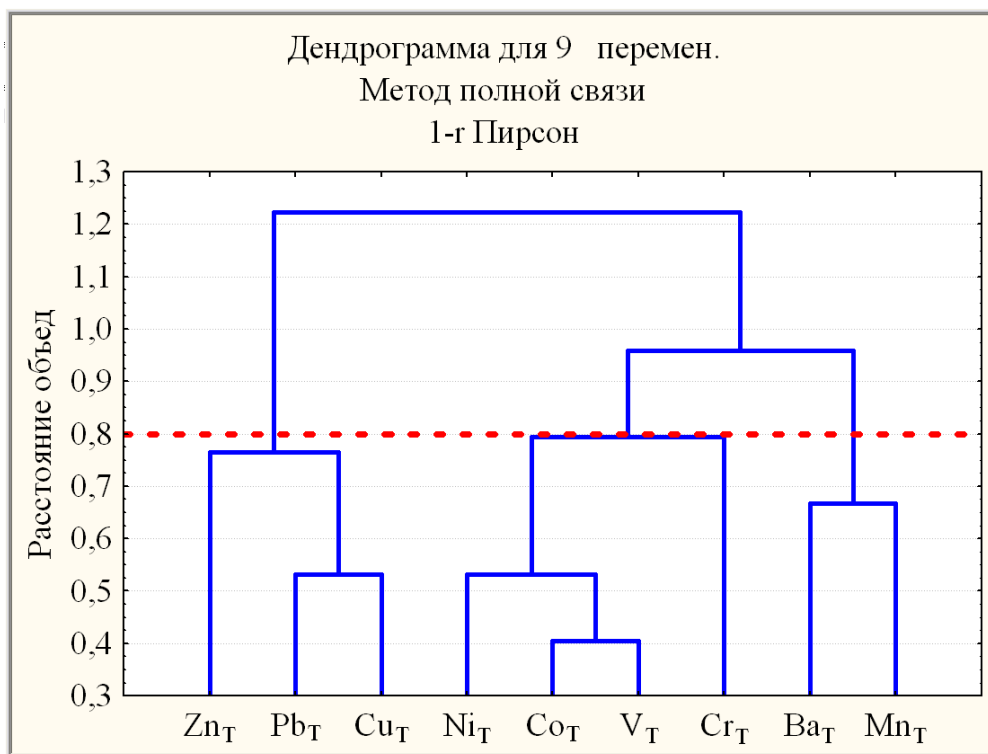


Рис. 3.41. Дендрограмма для М.
Метод взвеш. попарного среднего

Аналогичным образом строятся дендрограммы корреляционной матрицы микроэлементов в почве Т (рис.3.42 - 44).






Как следует из рис.3.39 -44, геохимические спектры микроэлементов в почве гг. Междуреченска и Томска образуют разные значимые микроассоциации химических элементов: {Cu – Co – Zn; V – Cr – Ni; Ba – Mn – Pb} и {Cu – Pb – Zn; V – Co – Ni – Cr; Ba – Mn; }.

3.1.7. Канонический анализ двух ассоциаций микроэлементов в почве территорий.

Если на предыдущем этапе анализировались парные зависимости микроэлементов, то в данном случае методом канонических корреляций анализируется зависимость между ассоциациями микроэлементов.

Рассмотрение проведем на примере содержаний химических элементов в почве территорий гг. Междуреченска (М) и Томска (Т), полученных методом ЭСП (файл Т – М ЭСП почва.sta на основе приложения П₂).

Для анализа составной зависимости рассматриваемых переменных вызывается диалоговое окно модуля  Канонический анализ :

Шаг 1. Запустите в полосе меню головной модуль  и в нем модуль  Канонический анализ (см. рис. 3.45).

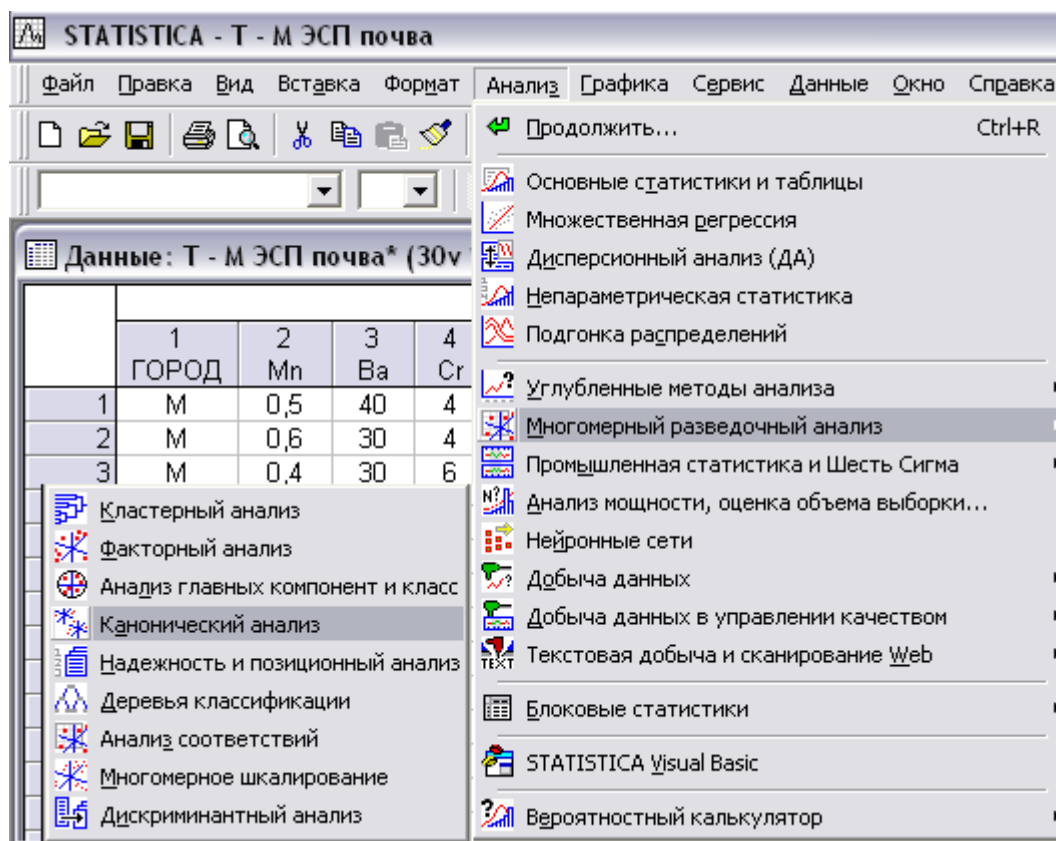

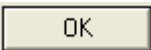


Рис. 3.45. Модуль  Канонический анализ в модуле 

Шаг 2. Выберите базы данных, как показано на рис. 3.46, и нажмите  в каждом окне.

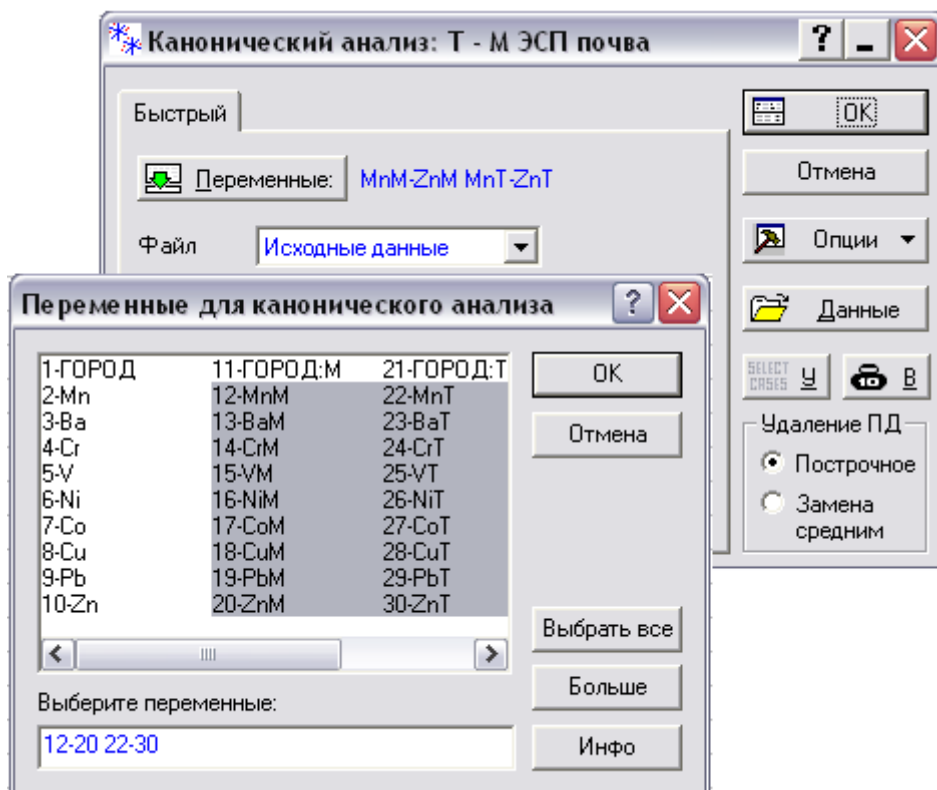


Рис. 3.46. Выбор исследуемых переменных для канонического анализа

На экране появится окно **Определение модели** (рис. 3.47).

Шаг 3. В окне **Определение модели** (рис. 3.47) можно просмотреть результаты вычисления выборочных числовых характеристик:

- ◇ **Средние и стандартные отклонения** - векторов математического ожидания и среднеквадратичного отклонения (рис. 3.48);
- ◇ **Корреляционная матрица** - матрицы парных коэффициентов корреляции (рис. 3.49);

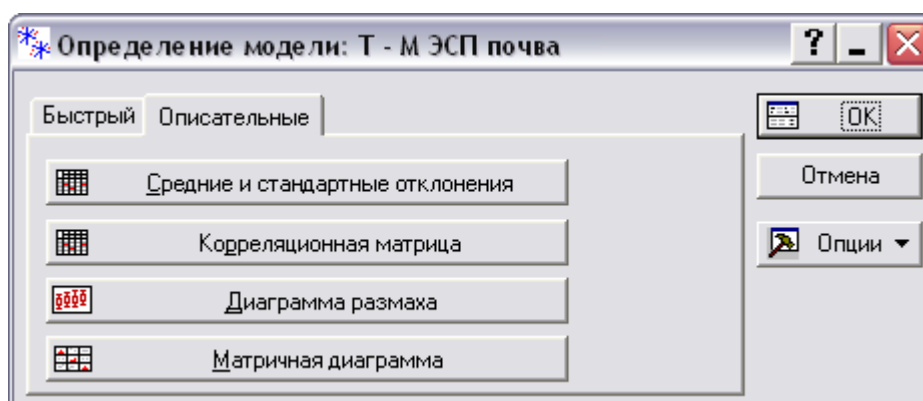




Рис. 3.47. Режим **Описательные** в окне **Определение модели**




Средние и станд. отклонения (Т - М ЭСП почва)					
	Средние		Ст. Откл.		
	Средние	Ст. Откл.	Средние	Ст. Откл.	
Mn _М	0,39300	0,19656	Mn _Т	0,42200	0,18726
Ba _М	40,80000	14,95313	Ba _Т	40,40000	15,56479
Cr _М	6,64000	2,63435	Cr _Т	11,02000	4,47209
V _М	5,65000	2,09557	V _Т	6,29000	2,20786
Ni _М	3,52000	1,12349	Ni _Т	3,33000	1,06415
Co _М	1,71300	0,92732	Co _Т	1,69500	0,86145
Cu _М	5,03000	1,80042	Cu _Т	5,57000	1,98100
Pb _М	3,35000	1,26631	Pb _Т	2,90000	1,28315
Zn _М	9,64000	3,72141	Zn _Т	6,68000	3,40789

Рис. 3.48. Векторы выборочных математического ожидания и средне-квадратичного отклонения

Корреляции (Т - М ЭСП почва)														
	Mn _М	Ba _М	Cr _М	V _М	Ni _М	Co _М	Cu _М	Pb _М	Zn _М	Mn _Т	Ba _Т	Cr _Т	V _Т	Ni _Т
Mn _М	1,00	0,49	0,30	0,27	0,18	0,30	0,19	0,46	0,20	-0,04	-0,03	-0,04	-0,05	0,08
Ba _М	0,49	1,00	0,32	0,27	-0,02	-0,09	-0,09	0,43	-0,06	-0,07	-0,04	0,14	-0,10	0,14
Cr _М	0,30	0,32	1,00	0,60	0,41	0,30	0,10	0,23	0,07	0,13	-0,05	0,18	-0,00	0,06
V _М	0,27	0,27	0,60	1,00	0,49	0,41	0,21	0,23	-0,01	0,09	0,16	0,19	0,07	-0,07
Ni _М	0,18	-0,02	0,41	0,49	1,00	0,44	0,10	0,24	0,13	0,01	0,10	0,08	0,12	-0,07
Co _М	0,30	-0,09	0,30	0,41	0,44	1,00	0,68	0,21	0,44	0,06	0,02	-0,04	0,06	-0,19
Cu _М	0,19	-0,09	0,10	0,21	0,10	0,68	1,00	0,22	0,42	0,17	0,02	-0,13	0,18	-0,01
Pb _М	0,46	0,43	0,23	0,23	0,24	0,21	0,22	1,00	0,37	-0,06	0,00	0,26	0,18	0,15
Zn _М	0,20	-0,06	0,07	-0,01	0,13	0,44	0,42	0,37	1,00	-0,03	0,05	0,00	0,10	-0,19
Mn _Т	-0,04	-0,07	0,13	0,09	0,01	0,06	0,17	-0,06	-0,03	1,00	0,33	0,22	0,36	0,31
Ba _Т	-0,03	-0,04	-0,05	0,16	0,10	0,02	0,02	0,00	0,05	0,33	1,00	0,13	0,21	0,12
Cr _Т	-0,04	0,14	0,18	0,19	0,08	-0,04	-0,13	0,26	0,00	0,22	0,13	1,00	0,45	0,37
V _Т	-0,05	-0,10	-0,00	0,07	0,12	0,06	0,18	0,18	0,10	0,36	0,21	0,45	1,00	0,47
Ni _Т	0,08	0,14	0,06	-0,07	-0,07	-0,19	-0,01	0,15	-0,19	0,31	0,12	0,37	0,47	1,00
Co _Т	0,05	-0,06	-0,12	-0,20	0,12	-0,02	0,08	0,26	0,09	0,16	0,04	0,21	0,59	0,47
Cu _Т	-0,04	-0,05	-0,09	-0,35	0,01	-0,15	-0,11	0,13	0,06	-0,06	0,02	-0,17	-0,04	0,10
Pb _Т	-0,13	-0,14	-0,06	-0,11	-0,06	-0,24	-0,08	-0,15	-0,17	0,07	0,13	-0,09	0,10	0,25
Zn _Т	0,12	0,02	0,15	0,12	-0,07	0,03	0,17	0,01	-0,08	0,16	0,14	-0,22	-0,04	0,14

Рис. 3.49. Матрица выборочных парных коэффициентов корреляции

Для визуализации корреляционной матрицы переменных можно построить диаграмму размахов  Диаграмма размаха и матрицу диаграмм рассеяния с гистограммами  Матричная диаграмма (рис. 3.50):

Шаг 4. В окне  Определение модели (рис. 3.47) нажмите кнопку  Матричная диаграмма, в открывшемся окне выберите некоторые переменные, например, Cr_M , V_M , Cu_T и Pb_T (аналогично рис. 3.46) и нажмите кнопку .

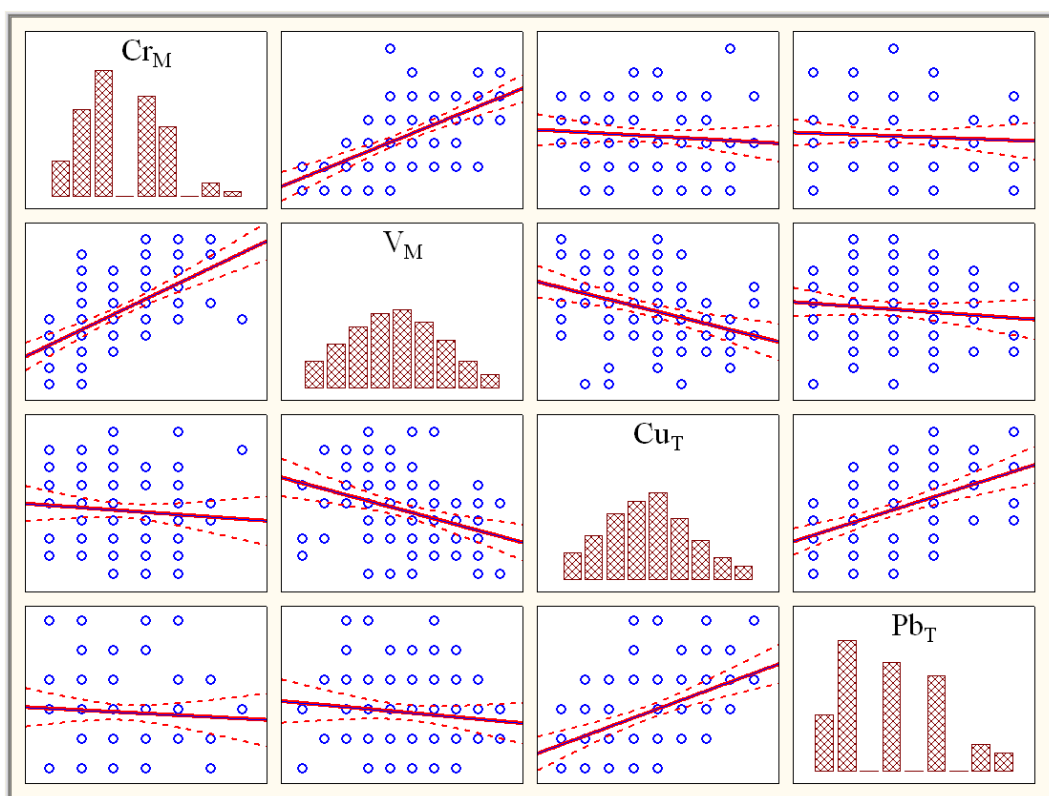

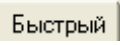
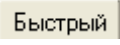





Рис. 3.50. Матрица гистограмм и диаграмм рассеяния с прямой линейной регрессии для исследуемых переменных

Подобные графики представляют наглядный и эффективный способ визуального анализа тесноты зависимостей между исследуемыми переменными, а гистограммы на диагонали для соответствующих переменных можно использовать для визуальной оценки закона нормального распределения переменной.

Для продолжения канонического анализа необходимо в окне  Определение модели (рис. 3.47) перейти в режим .

Шаг 5. В режиме  Быстрый окна  Определение модели определите микроэлементы М в первую группу, а микроэлементы Т - во вторую

группу (рис. 3.51) и нажмите , далее выделите галочкой **Пакетная обработка/печать** и нажмите . На экране появится таблица собственных чисел (рис. 3.52).

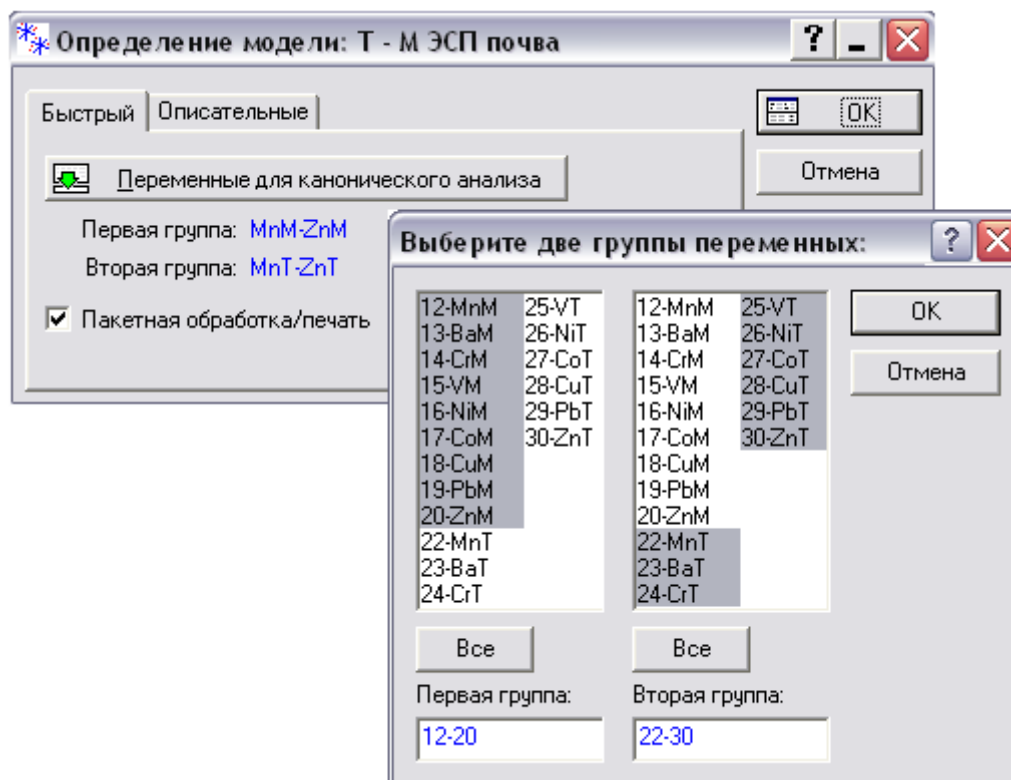


Рис. 3.51. Выбор двух групп переменных для канонического анализа

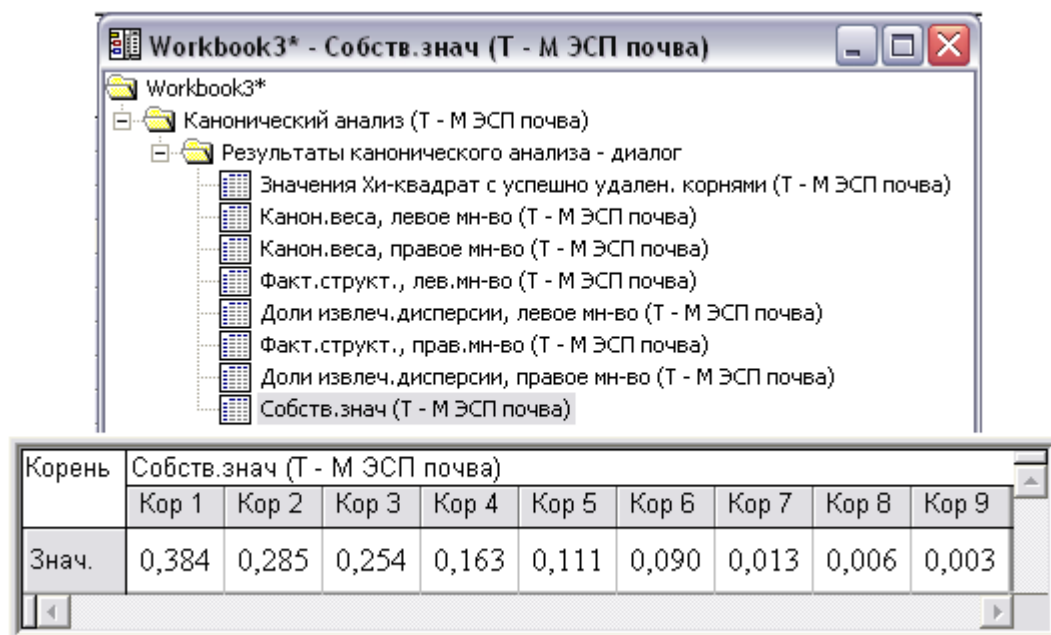



Рис. 3.52. Собственные значения λ_k^2 в режиме «Пакетная обработка»

При этом выше таблицы отражен столбцом протокол пакетного режима выполненных вычислений. Достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например,  Значения Хи-квадрат ... – для просмотра оценок значимости канонических корреляций R_k (рис. 3.53).

Корень Удаленный	Значения Хи-квадрат с успешно удален. корнями (Т - М ЭСП почва)				
	Канонич. R	Канонич. R-кв.	Хи-кв.	сс	p
0	0,620	0,384	136,438	81	0,0001
1	0,533	0,285	93,077	64	0,0103
2	0,504	0,254	63,110	49	0,0850
3	0,404	0,163	36,891	36	0,4276
4	0,333	0,111	20,960	25	0,6948
5	0,300	0,090	10,428	16	0,8433
6	0,114	0,013	1,988	9	0,9917
7	0,079	0,006	0,819	4	0,9358
8	0,053	0,003	0,256	1	0,6127

Рис. 3.53. Результаты оценки значимости коэффициентов канонических корреляций R_k ($R_k = \lambda_k$)



Согласно рис. 3.53, первому собственному значению $\lambda_1 \approx 0,384$ соответствует максимальный канонический коэффициент корреляции $R_1 \approx 0,62$. В следующем (третьем) столбце приводится расчетное значение статистики хи – квадрат $\chi^2 \approx 136,44$, которое при степени свободы $сс = 81$ указывает на высоко значимое ($p \approx 0,0001 < 0,0005$) отличие R_1 от нуля. Эта величина может быть проинтерпретирована как корреляция между первой парой U_1 и V_1 канонических переменных. Во второй строке приведена аналогичная оценка значимости коэффициента канонической корреляции $R_2 \approx 0,53$, для которого значение статистики хи – квадрат $\chi^2 \approx 93,08$ при степени свободы $сс = 64$ указывает статистически значимое ($0,005 < p \approx 0,01 < 0,05$) отличие R_2 от нуля. Эта величина может быть проинтерпретирована как корреляция между второй парой U_2 и V_2 канонических переменных.

Следует заметить, что последовательный критерий значимости работает следующим образом. Сначала рассматриваются все (9) канонические переменные вместе, т.е. без удаления корней. Полученное значение высоко значимо (на уровне значимости $p \approx 0,0001 < 0,0005$). Далее, первый (наиболее значимый) корень "удаляется" и определяется статистическая значимость 8-ми оставшихся корней. Это значение (во

второй строке таблицы результатов) статистически значимо ($0,005 < p \approx 0,01 < 0,05$). Затем "удаляются" два первых значимых корней и определяется статистическая значимость 7-ми оставшихся корней. Это значение (в третьей строке) слабо значимо ($0,05 < p \approx 0,085 < 0,10$). На этом можно остановиться и заключить, что только первые два корня является статистически значимым и должны быть изучены более подробно.

После вычисления собственных значений перейдем к построению соответствующих собственных векторов A_k и B_k (канонических весов). Канонические веса используются для вычисления значений канонических переменных. Для этого достаточно сложить исходные стандартизированные переменные с соответствующими весовыми коэффициентами:

$$U_k = \sum_{i=1}^9 a_{ki} X_i \text{ и } V_k = \sum_{j=1}^9 b_{kj} Y_j.$$

В режиме «Пакетная обработка» (рис. 3.52) достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например,  Канон.веса, левое мн-во ... — для просмотра коэффициентов a_{ki} (рис. 3.54) и  Канон.веса, правое мн-во ... — для просмотра b_{kj} (рис. 3.55)

Перем	Канон.веса, левое мн-во (Т - М ЭСП почва)								
	Корень 1	Корень 2	Корень 3	Корень 4	Корень 5	Корень 6	Корень 7	Корень 8	Корень 9
Mn _М	0,094	-0,438	-0,260	-0,127	0,400	0,561	-0,779	-0,432	0,310
Va _М	-0,124	0,047	0,225	0,330	0,536	0,110	1,075	0,258	0,396
Cr _М	-0,153	-0,586	0,535	-0,074	-0,214	-0,891	-0,323	-0,043	0,359
V _М	1,074	0,589	0,239	-0,166	-0,315	0,536	-0,021	-0,267	-0,085
Ni _М	-0,365	0,144	-0,633	0,021	-0,267	0,217	0,124	0,584	0,830
Co _М	-0,146	0,481	0,402	0,139	1,120	-0,445	-0,238	0,709	-0,749
Cu _М	0,256	-0,796	-0,629	-0,666	-0,383	-0,009	0,641	-0,012	0,265
Pb _М	-0,634	0,211	0,496	-0,701	-0,395	0,143	-0,158	0,072	-0,579
Zn _М	0,112	0,657	-0,216	0,139	0,059	-0,296	0,200	-0,833	0,495

Рис. 3.54. Вектор A_k канонических весов a_{ki} для $U_k = \sum_{i=1}^9 a_{ki} X_i$

Перем	Канон. веса, правое мн-во (Т - М ЭСП почва)								
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6	Корен 7	Корен 8	Корен 9
Mn _T	0,165	-0,348	-0,141	0,036	-0,037	-0,714	-0,106	0,599	0,536
Ba _T	0,136	0,460	-0,173	0,036	-0,038	0,587	0,168	-0,247	0,692
Cr _T	-0,114	0,486	0,925	-0,065	-0,343	-0,075	-0,455	0,203	0,040
V _T	0,226	0,240	-0,374	-0,528	-0,476	-0,506	0,837	-0,523	-0,559
Ni _T	-0,281	-0,778	0,337	0,129	0,028	0,528	0,734	-0,021	0,102
Co _T	-0,369	0,108	-0,434	-0,425	0,224	0,611	-0,946	0,685	0,118
Cu _T	-0,774	0,091	0,153	0,078	-0,085	-0,687	0,351	-0,621	0,311
Pb _T	0,336	-0,143	-0,278	0,629	-0,824	0,164	-0,448	0,154	-0,247
Zn _T	0,280	-0,200	0,393	-0,778	0,185	0,016	-0,475	-0,509	0,026

Рис. 3.55. Вектор V_k канонических весов b_{kj} для $V_k = \sum_{j=1}^9 b_{kj} Y_j$

Рассмотрим далее в режиме «Пакетная обработка» (рис. 3.52) обычные корреляции между каноническими переменными и переменными из каждого множества R_{UX} и R_{VY} (☐ Факт.структ., лев.мн-во ... – рис. 3.56, ☐ Факт.структ., прав.мн-во ... – рис. 3.57), которые называются каноническими нагрузками факторов и используются при интерпретации канонических переменных, так как считается, что переменные, сильно коррелированные с канонической переменной, имеют с ней много общего.

Перем	Факт.структ., лев.мн-во (Т - М ЭСП почва)								
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6	Корен 7	Корен 8	Корен 9
Mn _M	-0,058	-0,186	0,145	-0,409	0,558	0,404	-0,319	-0,206	0,394
Ba _M	-0,121	-0,115	0,588	-0,062	0,350	0,357	0,471	-0,030	0,387
Cr _M	0,178	-0,132	0,571	-0,277	-0,013	-0,402	-0,206	0,155	0,566
V _M	0,640	0,276	0,398	-0,392	-0,036	0,138	-0,080	0,260	0,328
Ni _M	-0,062	0,376	-0,148	-0,277	-0,081	0,017	-0,230	0,573	0,606
Co _M	0,215	0,269	-0,131	-0,551	0,565	-0,345	-0,123	0,328	0,021
Cu _M	0,263	-0,158	-0,367	-0,758	0,234	-0,267	0,263	0,033	-0,036
Pb _M	-0,451	0,240	0,369	-0,732	0,011	0,215	0,056	-0,107	0,074
Zn _M	-0,125	0,494	-0,228	-0,382	0,252	-0,381	0,083	-0,523	0,231

Рис. 3.56. Таблица парных корреляций $R_{U_k X_i}$ левого множества М

Перем	Факт.структ., прав.мн-во (Т - М ЭСП почва)								
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6	Корен 7	Корен 8	Корен 9
Mn _T	0,238	-0,276	-0,053	-0,273	-0,272	-0,397	0,097	0,443	0,594
Ba _T	0,245	0,326	-0,131	-0,094	-0,269	0,336	0,174	-0,162	0,752
Cr _T	-0,100	0,358	0,647	-0,228	-0,470	0,031	0,088	0,401	0,025
V _T	-0,039	0,121	-0,189	-0,637	-0,593	-0,020	0,371	0,191	-0,124
Ni _T	-0,271	-0,546	0,234	-0,272	-0,424	0,374	0,355	0,193	0,131
Co _T	-0,584	-0,046	-0,336	-0,523	-0,332	0,225	-0,170	0,283	0,022
Cu _T	-0,718	-0,117	-0,151	0,068	-0,262	-0,234	-0,211	-0,463	0,247
Pb _T	0,009	-0,344	-0,194	0,228	-0,752	0,100	-0,369	-0,281	0,031
Zn _T	0,266	-0,465	0,116	-0,419	-0,107	0,005	-0,415	-0,547	0,204

Рис. 3.57. Таблица парных корреляций $R_{V_k Y_j}$ правого множества Т

Согласно рис.3.56 , с канонической переменной U_1 наиболее сильно связана положительной корреляционной связью переменная V_M ($R \approx 0,64$), а согласно рис.3.57 , с V_1 наиболее сильно связана отрицательной корреляционной связью Cu_T ($R \approx -0,72$), что соответствует в перекрестной матрице x парных коэффициентов корреляции (рис. 3.49) самой сильной отрицательной корреляционной связи ($R \approx -0,35$) между химическими элементами V_M левого и Cu_T правого множеств, частотная диаграмма рассеяния которых (рис.3.58) построена аналогично рис.3.34-35.

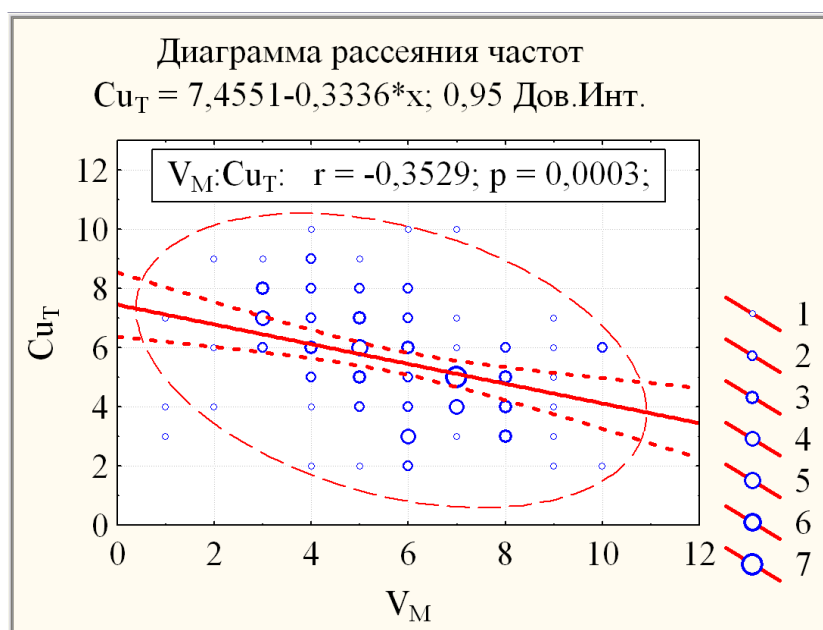


Рис. 3.58. Частотная диаграмма рассеяния $Cu_T : V_M$

Рассмотрим, наконец, в режиме «Пакетная обработка» (рис. 3.52) информацию об извлеченных дисперсиях и избыточности по группе микроэлементов М (Доли извлеч. дисперсии, левое мн-во ... — в случае левого множества) и группе микроэлементов Т (Доли извлеч. дисперсии, правое мн-во ... — в случае правого множества) для разных корней (рис. 3.59 и 3.60).

Workbook3* - Доли извлеч. дисперсии, левое мн-во (Т - М ЭСП почва)

Фактор	Доли извлеч. дисперсии, левое мн-во (Т - М ЭСП почва)	
	Извлечен дисперс.	Избыточн
Корень 1	0,088623	0,034030
Корень 2	0,075739	0,021550
Корень 3	0,134769	0,034224
Корень 4	0,226440	0,036922
Корень 5	0,097678	0,010844
Корень 6	0,094716	0,008524
Корень 7	0,057765	0,000749
Корень 8	0,095223	0,000597
Корень 9	0,129048	0,000369

Рис. 3.59. Таблица извлеченной дисперсии и соответствующей избыточности левого множества М для разных корней

Workbook3* - Доли извлеч. дисперсии, правое мн-во (Т - М ЭСП почва)

Фактор	Доли извлеч. дисперсии, правое мн-во (Т - М ЭСП почва)	
	Извлечен дисперс.	Избыточн
Корень 1	0,125426	0,048162
Корень 2	0,108109	0,030761
Корень 3	0,079535	0,020197
Корень 4	0,124566	0,020311
Корень 5	0,183900	0,020417
Корень 6	0,058503	0,005265
Корень 7	0,077060	0,000999
Корень 8	0,125562	0,000788
Корень 9	0,117339	0,000335

Рис. 3.60. Таблица извлеченной дисперсии и соответствующей избыточности правого множества Т для разных корней

Данная информация порождает суммарные характеристики канонического анализа (п. 1.2.5):

1) *Извлеченная дисперсия.* Коэффициенты канонической корреляции R_k ($k = \overline{1,9}$) соответствуют корреляции между каноническими переменными U_k и V_k , то есть между взвешенными суммами $U_k = \sum_{i=1}^9 a_{ki} X_i$ и

$V_k = \sum_{j=1}^9 b_{kj} Y_j$ по двум множествам переменных X и Y , соответствующими конкретному каноническому корню λ_k^2 . Они не говорят ничего о том, какую часть изменчивости (дисперсии) каждый канонический корень объясняет в переменных.

Однако, можно сделать заключение о суммарной по корням доле объясняемой дисперсии, рассматривая нагрузки канонических факторов, то есть корреляции R_{UX} и R_{VY} между каноническими переменными и исходными переменными в соответствующем множестве. Если возвести эти корреляции в квадрат, то полученные числа будут отражать долю дисперсии, объясняемую каждой переменной в случае каждого корня. Для каждого корня можно вычислить среднее значение этих долей и просуммировать по всем корням. При этом получится суммарная (по корням) средняя доля изменчивости объясненной в этом множестве на основании соответствующих канонических переменных. Другими словами, можно вычислить среднюю долю дисперсии, извлеченной суммой корней:

$$\frac{1}{9} \sum_{k=1}^9 \sum_{i=1}^9 R_{U_k X_i}^2 = 100\% \text{ для левого мн-ва и}$$

$$\frac{1}{9} \sum_{k=1}^9 \sum_{j=1}^9 R_{V_k Y_j}^2 = 100\% \text{ для правого мн-ва.}$$

Результат очевидный, так как все 9 корней извлекают 100% дисперсии как из левого множества переменных, так и из правого, поскольку программа извлекает в общем случае ровно столько корней, сколько переменных в меньшем множестве, то есть 9 в нашем случае. Если количества переменных в множествах разное, то только для одного из двух рассматриваемых множеств (с меньшим числом переменных) имеется столько канонических переменных, сколько исходных переменных

в нем содержится, что гарантирует 100% результат извлеченной дисперсии в этом множестве.

2) *Избыточность*. Коэффициент канонической корреляции R_k при возведении в квадрат дает долю дисперсии, общей для двух множеств. Если умножить эту долю на долю извлеченной дисперсии, то можно получить меру избыточности множества переменных, т.е., величину, показывающую, насколько избыточно одно множество переменных, если задано другое множество. Отметим, что можно вычислить избыточность первого (левого) множества переменных при заданном втором (правом) множестве, и избыточность второго (правого) множества переменных при заданном первом (левом) множестве. Поскольку последовательно извлекаемые канонические корни не коррелированы между собой, то можно просто просуммировать избыточности по всем (или только по значимым) корням, получив при этом общий коэффициент избыточности:

$$\frac{1}{9} \sum_{k=1}^9 \sum_{i=1}^9 R_{U_k X_i}^2 R_k^2 \approx 14,78 \% \quad \text{для левого мн-ва и}$$

$$\frac{1}{9} \sum_{k=1}^9 \sum_{j=1}^9 R_{V_k Y_j}^2 R_k^2 \approx 14,72 \% \quad \text{для правого мн-ва.}$$

Полученные значения избыточности можно объяснить следующим образом: пользуясь значениями всех канонических корней и получив значения переменных в правом множестве, можно объяснить, в среднем, 14,78 % дисперсии переменных в левом множестве. Аналогично, можно объяснить 14,72 % изменчивости в правой группе по значениям переменных в левом множестве. Эти результаты говорят о достаточно слабой зависимости между переменными двух множеств.

ВЫВОДЫ

❖ По совокупности критериев можно считать нормально распределенными в обоих городах содержания следующих микроэлементов: Mn, V и Cu (табл.3.3). Данные элементы мало специфичны для выбросов промышленных предприятий данных городов, что и подтверждается результатами.

❖ Как следует из табл. 3.7, высоко значимыми являются различия средних содержаний химических элементов Cr (6,64 и 11,02) и Zn (9,64 и 6,68); незначимыми являются различия средних содержаний Mn

(0,393 и 0,422); Ba (40,80 и 40,40); Ni (3,52 и 3,33) и Co (1,713 и 1,695); значимыми являются различия средних содержаний химических элементов V (5,65 и 6,29), Cu (5,03 и 5,57) и Pb (3,35 и 2,90). Для г. Междуреченска очень контрастно фиксируются в почвах ореолы загрязнения Pb и Zn. Это результат работы автостанции, а именно загрязнение от автотранспорта. Накопление Pb в почвах связано с работой двигателей автотранспорта долгое время на этилированном бензине, а присутствие Zn – это результат износа автошин. Контрастность ореолов данных элементов на фоне общего загрязнения от угольной промышленности (котельные, шахты, угольные разрезы), где данные элементы не специфичны.

❖ Значимые корреляционные связи между микроэлементами в почве г. Междуреченска лежат в пределах от 0,20 до 0,68, а в почве г. Томска в пределах от 0,21 до 0,59. Как следует из табл. 3.7 и 3.9, наиболее значимые корреляционные связи в ассоциациях наблюдаются в почве г. Междуреченска между Co и Cu, а в почве г. Томска между Co и V. Как следует из табл.3.10, из всех значимых корреляционных связей только три пары V & Co, Co & Cu и Cu & Zn значимо различаются по городам.

❖ Геохимические спектры микроэлементов в почве гг. Междуреченска и Томска образуют разные значимые микроассоциации химических элементов (рис.3.39 - 44): {Cu – Co – Zn; V – Cr – Ni; Ba – Mn – Pb} и {Cu – Pb – Zn; V – Co – Ni – Cr; Ba – Mn}. Результаты канонического анализа говорят о достаточно слабой зависимости между микроэлементами двух ассоциаций в почве территорий. В данном случае следует особо отметить тот факт, что для г. Томска характерно многопрофильность предприятий, что значительно сказывается на результаты исследований микроэлементов в почвах. В данном случае следует особо отметить предприятия с присутствием цехов с литейным производством, тогда как для г. Междуреченска – это избирательность производства и соответственно масштабы воздействия.

3.2. Статистический анализ (СА) эколого-геохимической информации по данным различных съемок

Проведем СА результатов содержаний ассоциации химических элементов, полученных методом ЭСП, территории района г. Стрежевого по данным снеговой (С), литогеохимической (П) и биогеохимической (З) съемок. С использованием STATISTICA 6.1 данный анализ может быть выполнен в следующей последовательности (аналогично п.2.1-4).

На предварительном этапе создается файл (с названием, напр., ЭГХ-съемки среды_П+С+З.sta) категориальной (СРЕДА_П+С+З) и раздельных по средам баз данных (Приложение П₃).

3.2.1. Вычисление выборочных характеристик содержаний химических элементов в ассоциациях по трем средам

В активизированном пакете STATISTICA 6.1 выполним последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль **Анализ**, выберите

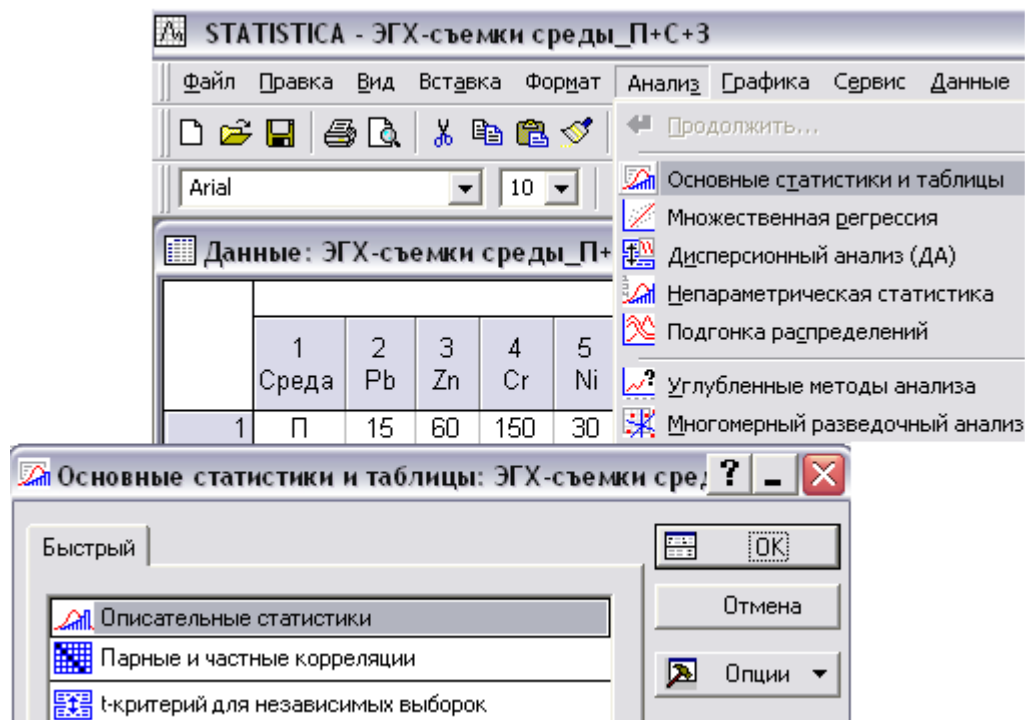
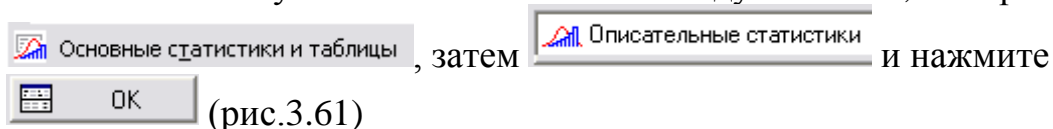

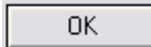
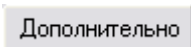
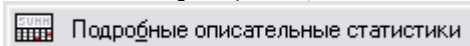



Рис.3.61. Открытие окна **Описательные статистики: ЭГХ-съемки среды_П+С+З** в модуле **Анализ**

Шаг 2. В открывшемся окне нажмите кнопку  и выберите переменные раздельной базы данных П (рис.3.62) и нажмите кнопку .

Шаг 3. Нажмите кнопку  и закажите (поставьте все нужные числовые характеристики (рис.3.62)): объем выборки (N), среднее (m), медиана (Me), мода (Mo), стандартное отклонение (S), стандартная ошибка среднего (σ_m), асимметрию (A), эксцесс (E) и их стандартные ошибки (σ_A , σ_E), а также минимум (Min) и максимум (Max).

Шаг 4. Нажмите кнопку  в окне  (рис.3.62).

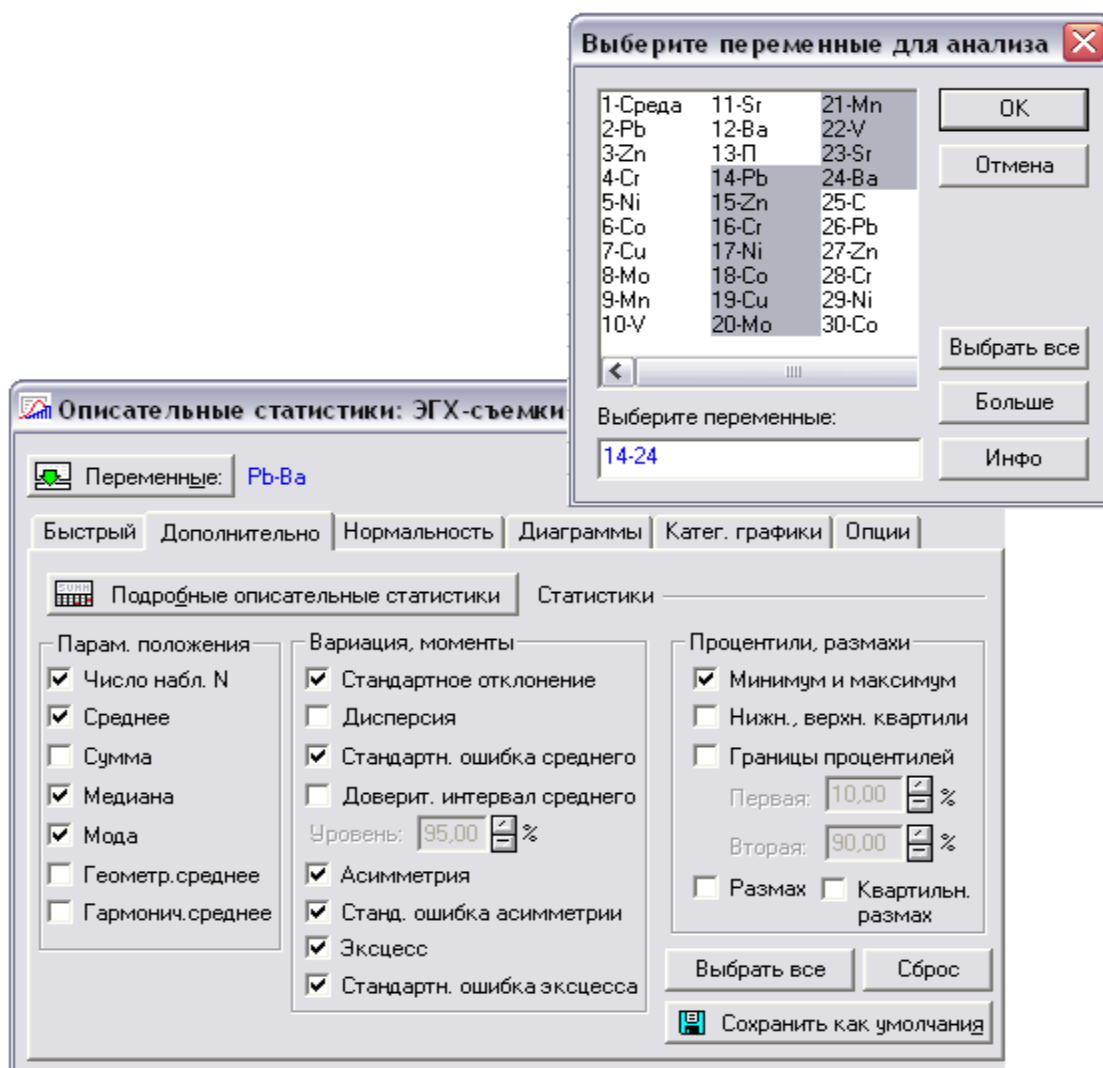


Рис.3.62. Режим  в окне .

Следующая таблица с результатами появится на экране:

Таблица 3.11

Числовые характеристики содержаний микроэлементов в почве (П) района г. Стрежевого.

(П)	<i>n</i>	<i>m</i>	<i>Me</i>	<i>Mo</i>	Min	Max	<i>S</i>	σ_m	<i>A</i>	σ_A	<i>E</i>	σ_E
Pb	33	22,58	16	11	9	71	14,66	2,55	1,76	0,41	3,21	0,80
Zn	33	73,64	60	40	40	300	52,01	9,05	2,96	0,41	10,84	0,80
Cr	33	159,4	150	150	60	300	58,68	10,21	1,01	0,41	1,11	0,80
Ni	33	29,58	30	30	6	60	11,48	2,00	0,68	0,41	1,81	0,80
Co	33	9,76	10	10	3	20	4,93	0,86	0,35	0,41	-0,72	0,80
Cu	33	18,30	15		6	60	9,83	1,71	2,52	0,41	9,44	0,80
Mo	33	2,70	3	3	2	4	0,59	0,10	0,15	0,41	-0,49	0,80
Mn	33	412,1	400	400	200	600	119,3	20,76	0,58	0,41	-0,77	0,80
V	33	33,03	30	30	10	60	8,47	1,47	0,34	0,41	3,25	0,80
Sr	33	133,3	100	100	100	200	47,87	8,33	0,74	0,41	-1,55	0,80
Ba	33	351,5	300	300	100	1000	158,4	27,57	2,18	0,41	8,34	0,80

Согласно табл. 3.11 содержания микроэлемента Cu не одно значение *Mo*, что свидетельствует о мультимодальном распределении (рис. 3.63, построенный аналогично рис.3.5-6).

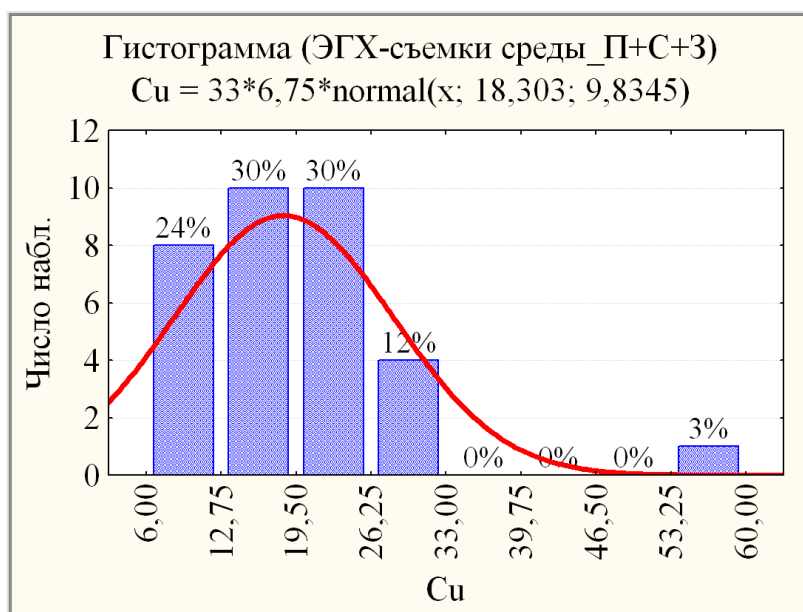


Рис. 3.63. Гистограмма распределения содержания Cu

Вычислим также числовые характеристики содержаний микроэлементов в С (табл.3.12) и З (табл.3.13).

Таблица 3.12

Числовые характеристики содержаний микроэлементов в снеговом (С) осадке района г. Стрежевого.

(С)	<i>n</i>	<i>m</i>	<i>Me</i>	<i>Mo</i>	<i>Min</i>	<i>Max</i>	<i>S</i>	σ_m	<i>A</i>	σ_A	<i>E</i>	σ_E
Pb	12	99,17	100,	100	40	200	43,16	12,46	0,99	0,64	2,09	1,23
Zn	12	82,50	60	60	40	150	426,0	123,0	2,82	0,64	8,16	1,23
Cr	12	69,17	60		40	150	29,99	8,66	1,84	0,64	4,73	1,23
Ni	12	19,25	12,5	10	8	40	12,45	3,60	0,81	0,64	-1,03	1,23
Co	12	5,17	5	3	3	8	2,08	0,60	0,32	0,64	-1,63	1,23
Cu	12	62,08	30		10	150	58,79	16,97	0,79	0,64	-1,27	1,23
Mo	12	3,58	3	3	2	6	1,31	0,38	0,94	0,64	0,31	1,23
Mn	12	325,0	300	300	100	600	160,3	46,26	0,47	0,64	-0,00	1,23
V	12	32,50	30		10	80	23,31	6,73	0,90	0,64	-0,19	1,23
Sr	12	258,3	300	300	200	300	51,49	14,86	-0,39	0,64	-2,26	1,23
Ba	12	391,7	300	300	300	800	156,4	45,16	2,04	0,64	3,91	1,23

Таблица 3.13

Числовые характеристики содержаний микроэлементов в золе (З) растен- ний района г. Стрежевого.

(З)	<i>n</i>	<i>m</i>	<i>Me</i>	<i>Mo</i>	<i>Min</i>	<i>Max</i>	<i>S</i>	σ_m	<i>A</i>	σ_A	<i>E</i>	σ_E
Pb	5	3,56	3,2		1,4	6,0	2,08	0,93	0,24	0,91	-2,73	2,00
Zn	5	140,8	117,3	117,3	117,3	176	32,15	14,38	0,61	0,91	-3,33	2,00
Cr	5	16,0	15,0		10,	20,0	4,18	1,87	-0,51	0,91	-0,61	2,00
Ni	5	13,56	12,0		4,8	24,0	7,56	3,38	0,44	0,91	-0,89	2,00
Co	5	1,60	1,6	1,6	0,6	3,0	0,88	0,39	1,02	0,91	2,00	2,00
Cu	5	12,76	11,0		8,8	16,5	3,53	1,58	0,30	0,91	-2,72	2,00
Mo	5	2,60	3,0	3,0	2,0	3,0	0,55	0,24	-0,61	0,91	-3,33	2,00
Mn	5	448,0	420		280	560	117,1	52,38	-0,51	0,91	-0,61	2,00
V	5	3,00	3,0		1,0	4,0	1,22	0,55	-1,36	0,91	2,00	2,00
Sr	5	336,0	240	240	240	480	131,5	58,79	0,61	0,91	-3,33	2,00
Ba	5	224,0	210		140	280	58,57	26,19	-0,51	0,91	-0,61	2,00

3.2.2. Однофакторный дисперсионный анализ (п.2.3.1) содержания химических элементов в ассоциациях по трем средам

Для проведения оценки значимости различия содержания химических элементов по трем средам (почва, снег и зола) воспользуемся сначала средствами параметрического однофакторного дисперсионного анализа применительно к категориальной базе данных в файле с названием ЭГХ-съемки среды П+С+3.sta (Приложение П₃).

Шаг 1. Запустите **Анализ** и щелкните **Дисперсионный анализ (ДА)** (см. рис. 3.64). В появившемся окне выберите вид анализа **Однофакторный ДА** и задайте метод (диалог быстрых спецификаций), как показано на рис. 3.64, и нажмите **ОК**.

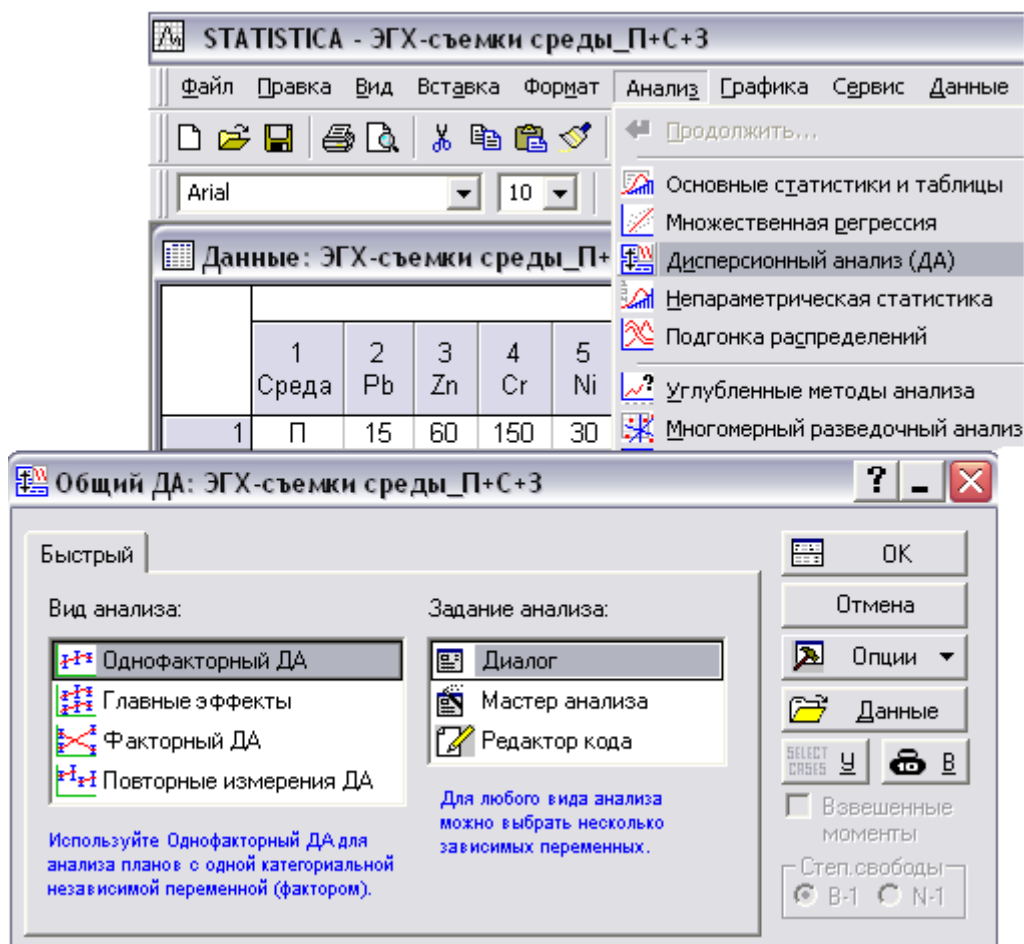


Рис. 3.64. Модуль **Дисперсионный анализ (ДА)** в головном модуле **Анализ** для базы данных **ЭГХ-съемки среды П+С+3**

Шаг 2. В открывшемся окне нажмите кнопку Переменные: и выберите переменные: в качестве зависимых - содержания химических элементов (2–12) , а в качестве категориальной – СРЕДА (рис. 3.65) и нажмите . Далее нажмите кнопку Коды факторов: , выберите в окне и нажмите . После этого нажмите кнопку в .

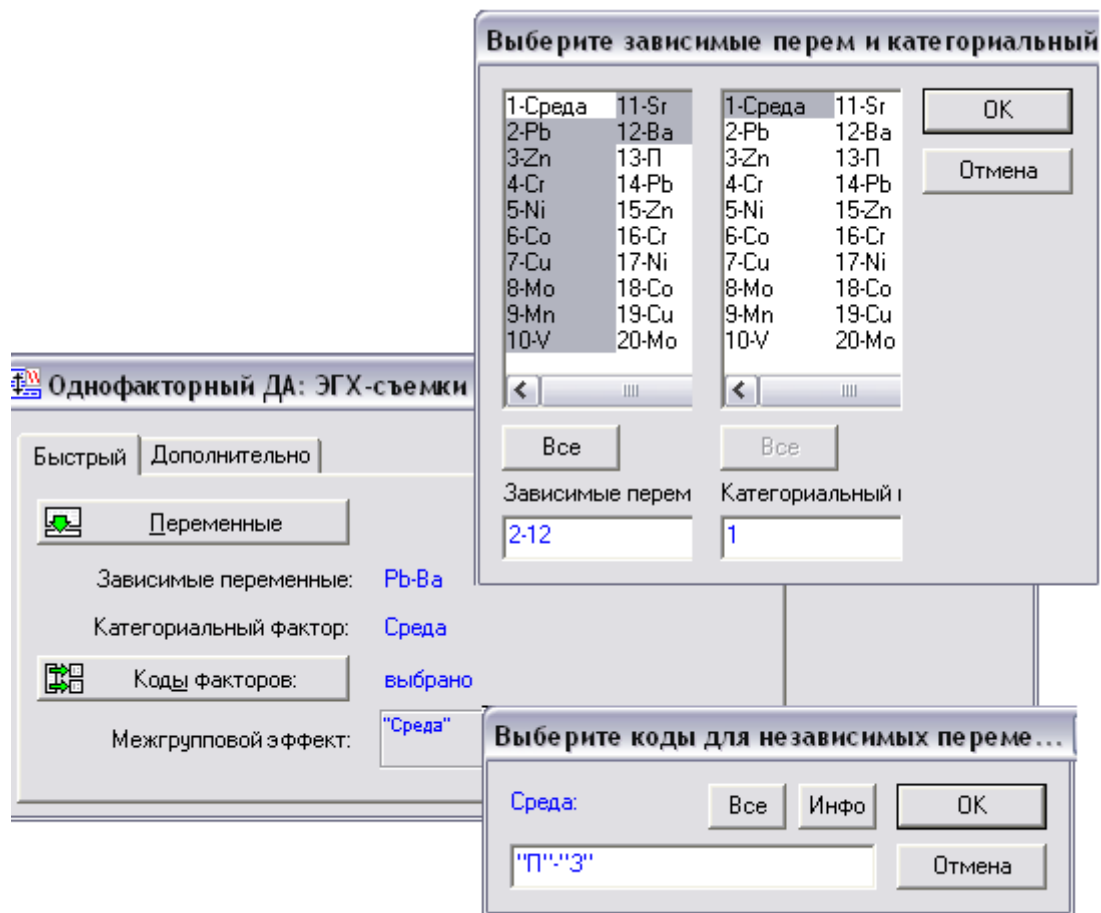


Рис. 3.65. Выбор переменных для

Шаг 3. В появившемся окне Результаты анализа (рис. 3.66) в режиме Итоги нажмите Статистики ячеек для просмотра результатов описательной статистики (аналогично табл. 3.11-13) в усеченном формате (рис. 3.67) для всех химических элементов во всех средах и Одномерн. результаты для просмотра в стандартном формате результатов дисперсионного анализа различий содержаний каждого химического элемента по совокупности сред (например, для Pb и Zn, рис. 3.68).

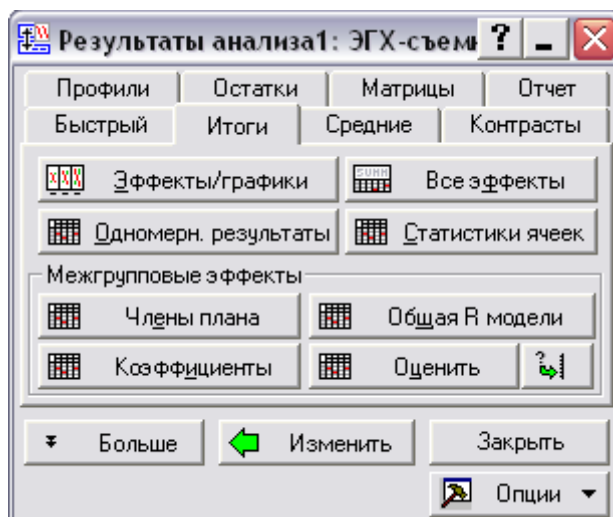



Рис. 3.66. Окно  Результаты анализа1: ЭГХ-съем

Эффект	Описательные статистики (ЭГХ-съемки среды_П+С+З)							
	Уровень Фактор	N	Pb Среднее	Pb Ст.откл.	Pb Стд.ош.	Pb -95,00%	Pb +95,00%	Zn Среднее
Всего		50	39,06	41,90	5,92	27,15	50,96	82,48
Среда	П	33	22,58	14,66	2,55	17,38	27,77	73,64
Среда	С	12	99,17	43,16	12,46	71,74	126,59	82,50
Среда	З	5	3,56	2,08	0,93	0,98	6,14	140,78

Рис. 3.67. Результаты описательной статистики ЭГХ-съемки


На рис. 3.67 кроме общей и групповых средних приведены также соответствующие стандартные отклонения, стандартные ошибки и границы 95% - ых доверительных интервалов.





Эффект	Одномерные результаты для каждой ЗП (ЭГХ-съемки среды_П+С+З) Сигма-ограниченная параметризация Декомпозиция гипотезы									
	Степени свободы	Pb SS	Pb MS	Pb F	Pb p	Zn SS	Zn MS	Zn F	Zn p	
Среда	2	58622,1	29311,0	50,3	0,0000	19575,4	9787,7	4,1	0,0228	
Ошибка	47	27385,0	582,7			112123,5	2385,6			
Всего	49	86007,0				131698,8				

Рис. 3.68. Результаты дисперсионного анализа различий содержаний химических элементов по совокупности сред

На рис. 3.68 приведена таблица результатов дисперсионного анализа в стандартном формате. В столбце SS приведены значения сумм квадратов отклонений: например, для Zn $SS_B = 19575,4$ – межгрупповой, $SS_R = 112123,5$ – внутригрупповой и $SS_T = 131698,8$ – общей вариацией переменной, складывающейся в силу (1.30) из двух компонент SS_B и SS_R . В первом столбце приведены соответствующие степени свободы: $\nu_B = 2$, $\nu_R = 47$ и $\nu_T = 49$. В столбце MS приведены значения средних квадратов $MS_B = 9787,7$ и $MS_R = 2385,6$, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы. В столбце F приведено значение наблюдаемого F-отношения, вычисленное по формуле $F_{набл} = MS_B / MS_R \approx 4,1$, а в следующем столбце – соответствующее значение уровня значимости $p_F \approx 0,023$, определяющее различие содержаний Zn по совокупности сред как статистически значимое ($0,005 < p < 0,05$).

Согласно рис. 3.48 высоко значимыми по совокупности трех сред являются различия содержаний следующих химических элементов Pb, Cr, Co и V; незначимыми по совокупности трех сред являются различия средних содержаний Mn и Ba. Остальные химические элементы имеют промежуточные варианты различия содержаний по средам.

Продемонстрируем возможности модуля  Дисперсионный анализ (ДА) для расширенного анализа полученного результата.

Шаг 3'. Нажмите  Эффекты/графики (рис. 3.66). После этого в появившемся окне  Таблица всех эффектов выделить опцию  График и два раза щелкнуть на строке результатов для просмотра графика средних выбранной (рис. 3.69) переменной, например, Zn с усами 95% - ых доверительных интервалов (рис.3.70) или  Таблицу - для просмотра описательной статистики (рис. 3.67)

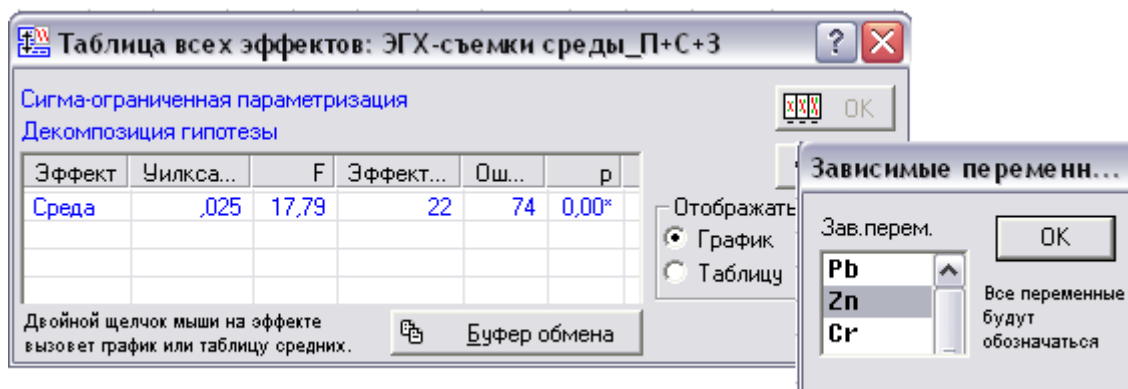


Рис. 3.69. Диалоговое окно  Таблица всех эффектов (содержит результаты анализа) для просмотра таблиц и графика средних

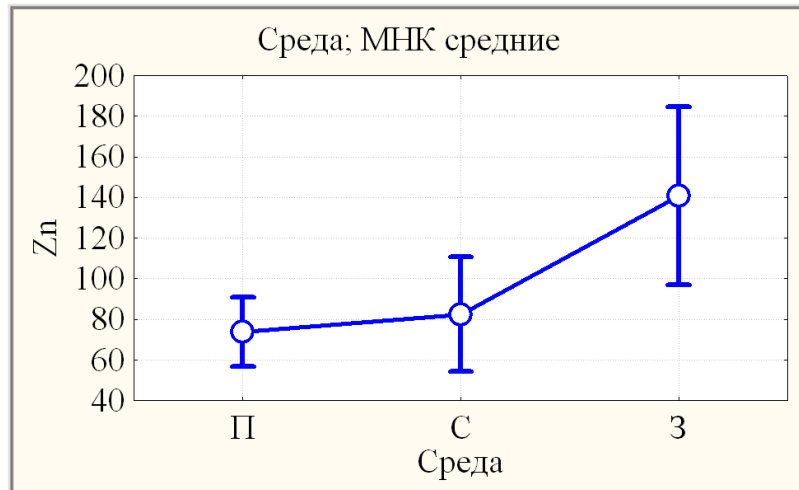


Рис. 3.70. График групповых средних переменной Zn с усами 95% доверительных интервалов

Так как, согласно результатам дисперсионного анализа (рис.3.69), существуют значимые различия между средними в группах по переменным, то далее можно использовать метод множественных сравнений.

Шаг 4. В окне **Результаты анализа** (рис. 3.66) нажмите кнопку **Больше** и в открывшемся окне выберите режим **Апостериорные**, в котором представлены различные апостериорные критерии (рис. 3.71). Выберите последовательно все критерии **Фишера НЗР**, **Бонферрони**, **Шеффе**, **Тьюки** для переменной Zn (рис. 3.72).

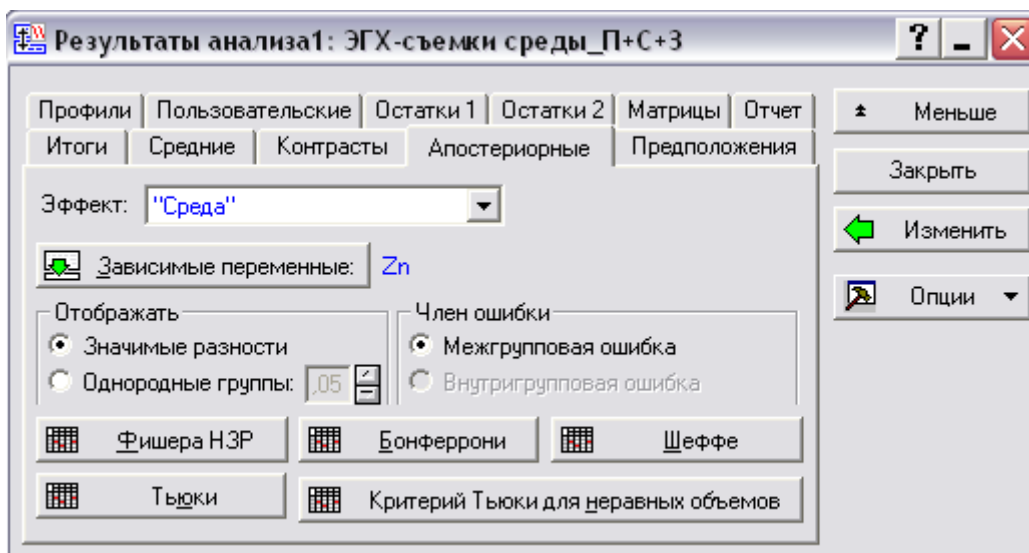


Рис. 3.71. Режим **Апостериорные** для выбора апостериорных критериев множественных сравнений средних

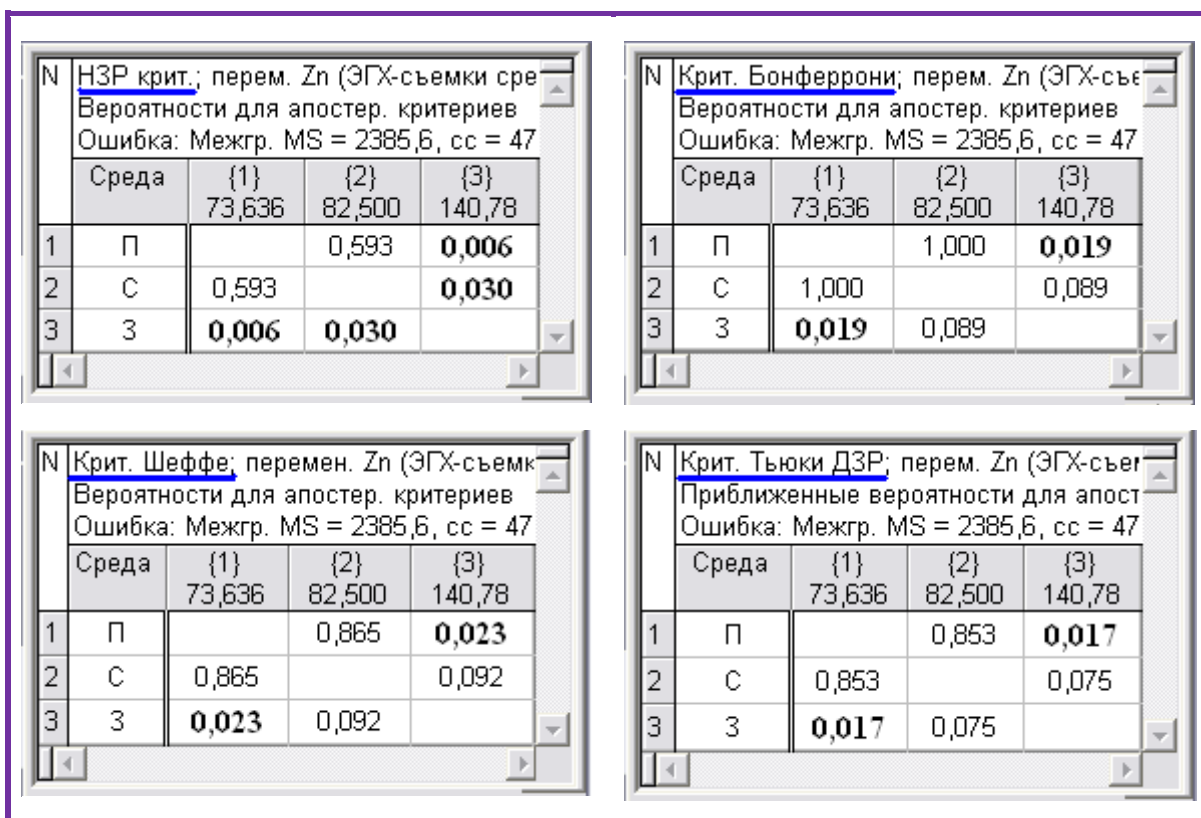


Рис. 3.72. Результаты сравнений средних групп Zn разными апостериорными критериями

Согласно рис. 3.72, в случае с Zn все апостериорные критерии единодушны в оценке уровней значимостей $p_{П-С}$ и $p_{П-З}$: средние П и С различаются незначимо ($0,10 < p_{П-С}$), а средние П и З различаются статистически значимо ($0,005 < p_{П-З} < 0,05$). При этом, среднее З отличается от С статистически значимо ($0,005 < p_{С-З} < 0,05$) по критерию

Фишера НЗР

и слабо значимо ($0,05 < p_{С-З} < 0,10$) по критериям

Бонферрони,

Шеффе,

Тьюки.

Различия в результатах применения разных критериев объясняются соответствующими особенностями и рекомендациями по их применению. Например, метод Шеффе допускает различные объемы выборок, в то время как метод Тьюки применяется лишь при равных объемах выборок. Множественный t -метод применяется только к множеству контрастов, выбранных до начала исследования данных, в то время как в двух других методах множество контрастов может быть любым.

Полученные таким образом (рис. 3.68 и 3.72 по апостериорному критерию Фишера НЗР) результаты сравнений групповых средних каждого химического элемента сведены в следующую таблицу:


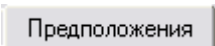
Таблица 3.14

Результаты параметрического однофакторного дисперсионного анализа ЭГХ-съемки среды

	p_F	$p_{П-С}$	$p_{П-З}$	$p_{С-З}$	$m_{П}$	$m_{С}$	$m_{З}$
Pb	0,00000	0,00000	0,10736	0,00000	22,58	99,17	3,56
Zn	0,02279	0,59288	0,00622	0,02974	73,64	82,50	140,8
Cr	0,00000	0,00000	0,00000	0,05410	159,39	69,17	16,00
Ni	0,00297	0,01021	0,00541	0,35502	29,58	19,25	13,56
Co	0,00009	0,00218	0,00019	0,11728	9,76	5,17	1,60
Cu	0,00018	0,00006	0,69810	0,00299	18,30	62,08	12,76
Mo	0,00656	0,00224	0,80486	0,02774	2,70	3,58	2,60
Mn	0,09753	0,05240	0,56754	0,08163	412,12	325,0	448,0
V	0,00009	0,90615	0,00002	0,00013	33,03	32,50	3,00
Sr	0,00000	0,00000	0,00000	0,01970	133,33	258,3	336,0
Ba	0,12563	0,43708	0,08689	0,04369	351,52	391,7	224,0

Как следует из табл. 3.14, картина значимости различий содержания элементов по средам является весьма неоднородной. Даже в случае высоко значимого ($p_F < 0,0005$) различия по совокупности трех сред для **Pb** ($m_C = 99,17$; $m_{П} = 22,58$ и $m_З = 3,56$ за счет $p_{П-С}$ и $p_{С-З}$) допускается незначимое различие ($0,10 < p_{П-З}$) средних $m_{П} = 22,58$ и $m_З = 3,56$. С другой стороны, незначимое различие ($0,10 < p_F$) по совокупности трех сред для **Ba** ($m_C = 391,7$; $m_{П} = 351,5$ и $m_З = 224,0$) допускается статистически значимое различие ($0,005 < p_{С-З} < 0,05$) средних $m_C = 391,7$ и $m_З = 224,0$.

Чтобы убедиться в достоверности полученных результатов параметрического дисперсионного анализа, необходимо проверить предположения о нормальном распределении сравниваемых групп и об однородности дисперсий в группах.

Шаг 5. В окне  (рис. 3.71) выберите режим  (рис. 3.73), в котором представлены различные критерии проверки гипотезы однородности дисперсий и графические средства проверки соответствия закона распределения переменной нормальному закону. Выберите, например, критерий однородности дисперсий

Критерий Левена (рис. 3.74) и Гистограммы распределения переменной для всех групп (например, для Zn, рис. 3.75).

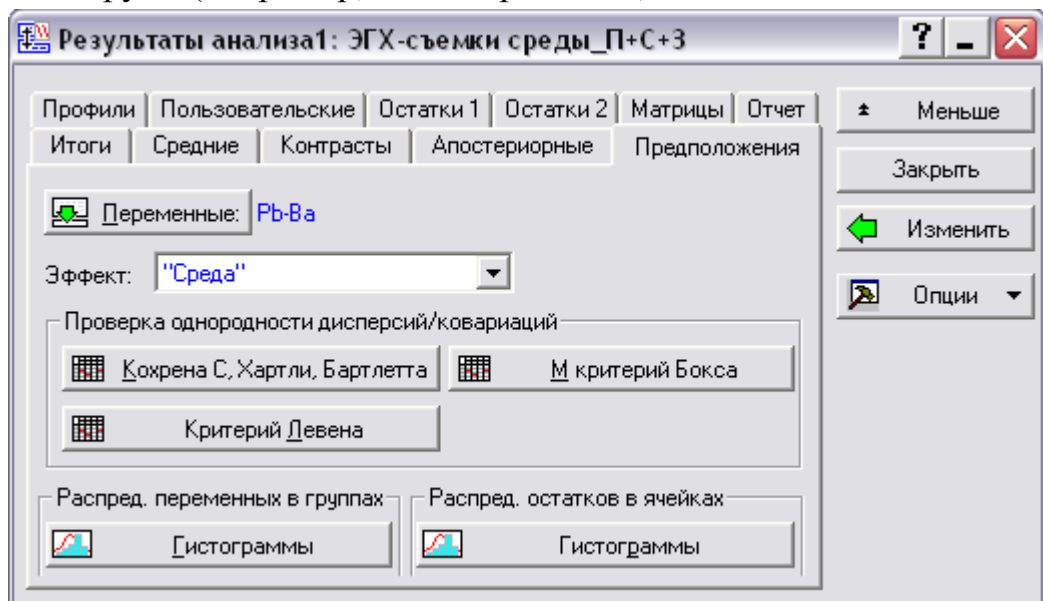


Рис. 3.73. Режим Предположения для проверки предположений дисперсионного анализа

Критерий Левена однородности дисперсий (ЭГХ-съемки среды_П+С+З)
Эффект: "Среда"
Степени свободы для всех F: 2, 47

	MS Эффект	MS Ошибка	F	p
Pb	1403,70	323,30	4,34176	0,018609
Zn	129,35	1167,40	0,11080	0,895351
Cr	4703,59	1160,26	4,05390	0,023757
Ni	60,24	61,12	0,98564	0,380780
Co	36,72	5,98	6,14141	0,004270
Cu	8987,72	200,87	44,74459	0,000000
Mo	1,19	0,19	6,13346	0,004297
Mn	3033,18	6706,96	0,45224	0,638938
V	809,24	62,54	12,93884	0,000033
Sr	10989,67	249,43	44,05944	0,000000
Ba	7774,68	12689,68	0,61268	0,546169

Рис. 3.74. Результаты проверки гипотезы об однородности дисперсий с помощью Критерий Левена

Согласно рис. 3.74 дисперсии различаются незначимо ($0,10 < p$) в распределениях по средам только для химических элементов Zn, Ni, Mn и Ba.

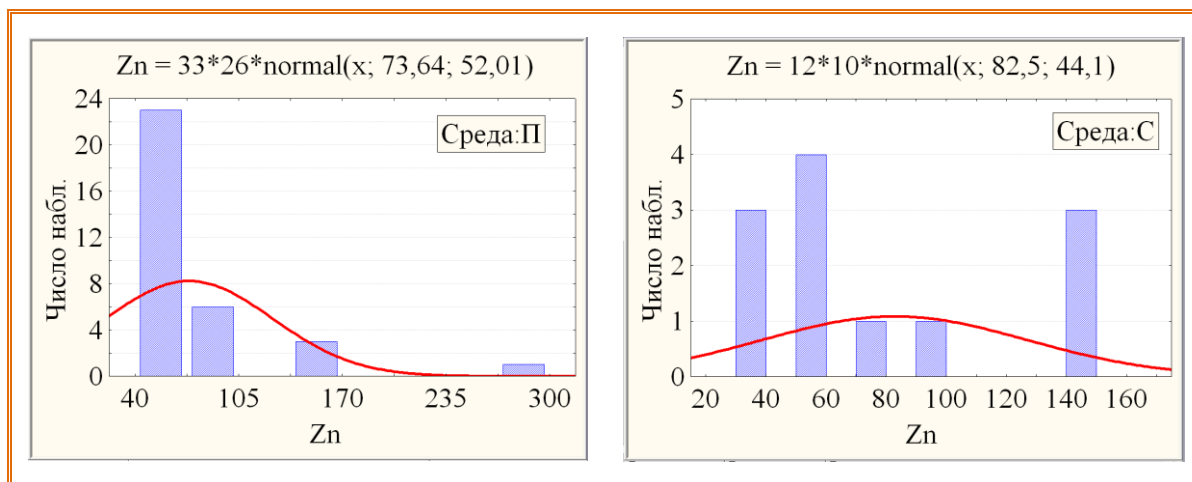


Рис. 3.75. Гистограммы распределения переменной Zn для сред П и С.

Согласно рис. 3.75 распределения химического элемента Zn по средам зримо отличаются от нормального закона.

В связи с нарушением предположений о нормальном распределении сравниваемых групп, соответствующих уровням фактора СРЕДА, и об однородности дисперсий в группах обратимся к непараметрической (ранговой) альтернативе однофакторного дисперсионного анализа: тесту Краскела-Уоллиса.

Шаг 6. Запустите модуль **Анализ** и щелкните в нем по кнопке **Непараметрическая статистика** (см. рис.3.64). Затем выберите сравнение нескольких независимых групп **Сравнение нескольких независимых групп** (рис.3.76).

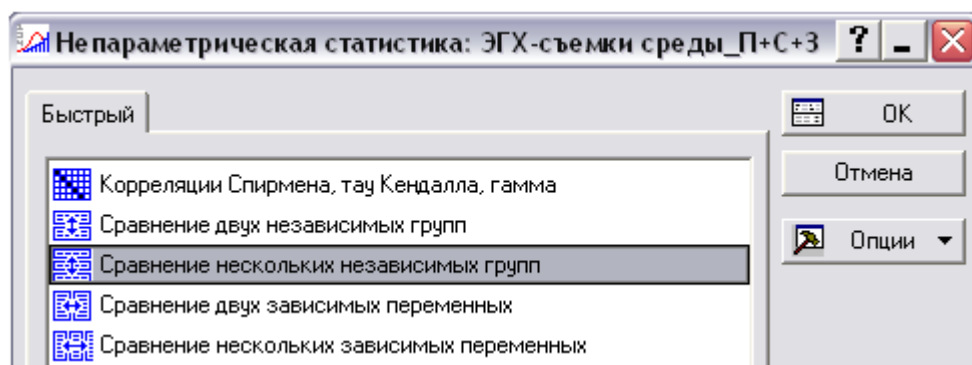




Рис. 3.76. Выбор **Сравнение нескольких независимых групп** в окне

Непараметрическая статистика:

Далее в  **ДА Краскела-Уоллиса** закажите переменные стандартным образом (см. рис. 3.77) и щелкните по кнопке  **ДА Краскела-Уоллиса** для вывода на экран результатов рангового однофакторного дисперсионного анализа: по тесту Краскела-Уоллиса (например, для Zn, рис. 3.78).

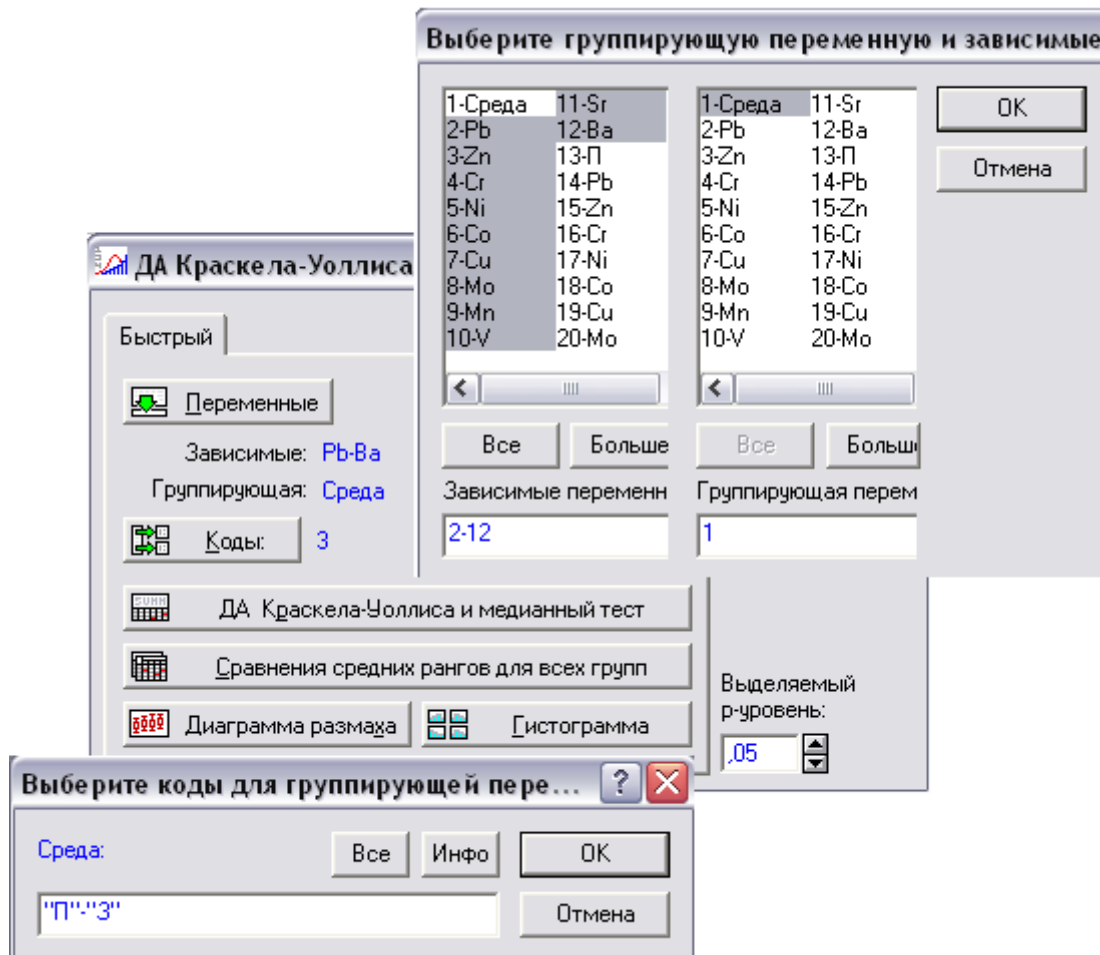



Рис.3.77. Диалоговое окно  **ДА Краскела-Уоллиса**

Зависим.: Zn	Ранговый ДА Краскела-Уоллиса; Zn (ЭГХ-съемки среды_П+С+З) Группирующая переменная: Среда Кр.Краскела-Уоллиса: $H(2, N=50) = 9,629609$ $p = .0081$		
	Код	Допуст N	Сумма Ряды
П	101	33	739,5
С	102	12	318,5
З	103	5	217,0

Рис.3.78. Результаты рангового дисперсионного анализа СРЕДА по тесту Краскела-Уоллиса

Согласно критерию Краскела-Уоллиса (рис. 3.78) существует статистически значимое различие ($0,005 < p_{к.у} \approx 0,008 < 0,05$) распределений химического элемента Zn по совокупности сред. Поэтому далее можно использовать непараметрический аналог метода множественных сравнений.

Шаг 7. В окне **ДА Краскела-Уоллиса** (рис. 3.78) нажмите кнопку



 Сравнения средних рангов для всех групп (рис. 3.79).

Зависим.: Zn	Сравнения p значений (2-стороннее); Zn (ЭГХ-съемки среды)		
	Группирующая переменная: Среда		
	Кр.Краскела-Уоллиса: $H(2, N=50) = 9,629609$ ($p = ,0081$)		
	П R:22,409	С R:26,542	З R:43,400
П		1,000000	0,00808
С	1,000000		0,089424
З	0,00808	0,089424	

Рис. 3.79 Результаты сравнений средних рангов сред Zn

Согласно рис. 3.79, средние П и С различаются незначимо ($0,10 < p_{п.с} < 0,05$), среднее З отличается от П статистически значимо ($0,005 < p_{п.з} < 0,05$), а от С - слабо значимо ($0,05 < p_{с.з} < 0,10$).

Шаг 7'. В окне **ДА Краскела-Уоллиса** (рис.3.77) нажмите кнопку

 Диаграмма размаха для просмотра диаграммы размаха по группам (например, для Zn, рис. 3.80), и –  Гистограмма для просмотра гистограмм по группам (рис.3.75)

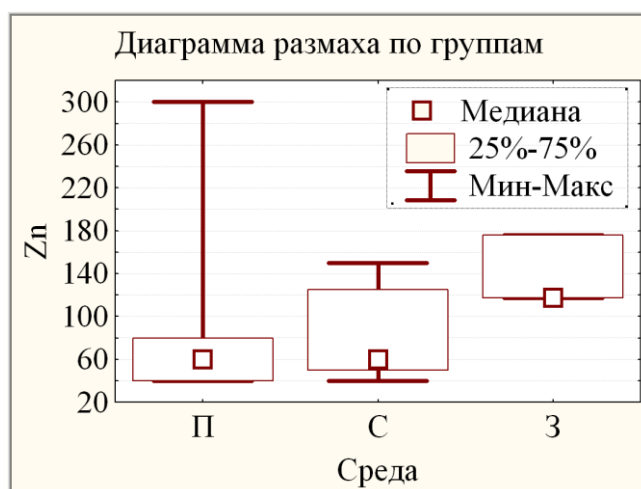


Рис. 3.80. Диаграммы размаха Zn по группам

Полученные таким образом (рис. 3.78 и 3.79) результаты сравнений групповых средних каждого химического элемента сведены в следующую таблицу:

Таблица 3.15

Результаты непараметрического однофакторного дисперсионного анализа ЭГХ-съемки среды

	p_{K-Y}	$p_{П-С}$	$p_{П-З}$	$p_{С-З}$	Me_{II}	Me_C	Me_3
Pb	0,0000	0,0000	0,0181	0,0000	16,0	100	3,2
Zn	0,0081	1,0000	0,0081	0,0894	60,0	60,0	117,3
Cr	0,0000	0,0001	0,0000	0,5489	150	60,0	15,0
Ni	0,0027	0,0582	0,0167	0,9243	30,0	12,5	12,0
Co	0,0000	0,0228	0,0002	0,1854	10,0	5,0	1,6
Cu	0,0223	0,0813	0,7796	0,0472	15,0	30,0	11,0
Mo	0,0450	0,0971	1,0000	0,3417	3,0	3,0	3,0
Mn	0,0997	0,2127	1,0000	0,2451	400	300	420
V	0,0005	1,0000	0,0006	0,0128	30,0	30,0	3,0
Sr	0,00000	0,0001	0,0006	1,0000	100	300	240
Ba	0,0069	1,0000	0,0178	0,0113	300	300	210

Как следует из табл. 3.15, картина значимости неоднородностей содержаний элементов по средам является весьма пестрой. Даже в случае высоко значимого ($p_{K-Y} < 0,0005$) различия по совокупности трех сред, например, для **Cr** и **Sr** (за счет $p_{П-С}$ и $p_{П-З}$) допускается незначимое различие ($0,10 < p_{С-З}$) сред С и З.

В случае расхождения результатов рангового (табл. 3.15) и параметрического (табл. 3.14) дисперсионных анализов в связи с нарушением условий применения последнего более корректными признаются результаты непараметрического дисперсионного анализа. Так, например, в случае **Ba** ($0,10 < 0,13 \approx p_F$ и $0,005 < p_{K-Y} \approx 0,007 < 0,05$) следует признать статистически значимое различие по совокупности трех сред.

Наглядная демонстрация значимости различия содержания химического элемента по средам визуально реализуется построением составной гистограммы. Для построения составных по средам II, С и З гистограмм распределений содержаний химических элементов необходимо выполнить последовательность следующих действий (п. 2.1.3):

Шаг 1. Запустите в головном меню модуль **Графика**, выберите **Категоризованные графики** в стартовой панели модуля и далее щелкните мышкой на разновидности **Гистограммы...** (рис.3.81).

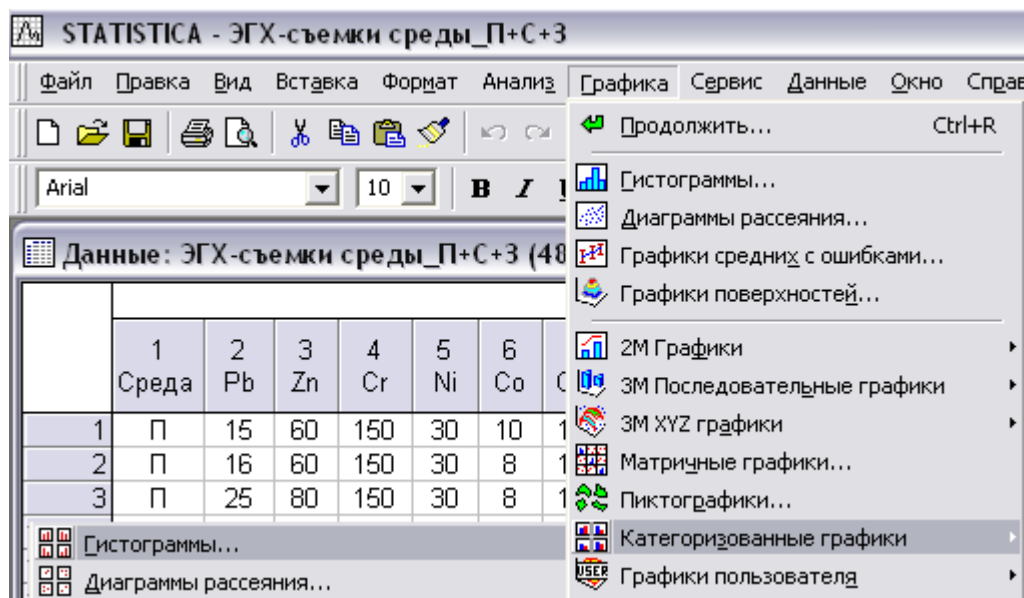


Рис.3.81. Открытие окна **2М Категоризованные гистограммы** в модуле **Графика**

Шаг 2. В появившемся окне **2М Категоризованные гистограммы** щелкните **Дополнительно** (рис.3.82).

Шаг 3. В открывшемся окне (рис.3.82) нажмите кнопку **Переменные:** и выберите переменные категориальной базы данных (например, переменная «Ni» – в окне **Переменная:** и «Среда» – в окне **X-Категория:**) и нажмите ОК.

Шаг 4. Далее в окне **X-Категории:** выделите режим **Коды:**, нажмите кнопку **Задать коды**, выберите **Все** в **Коды категорий:** (рис. 3.82) и нажмите ОК.

Шаг 5. После этого в окне **Размещение:** выделите режим **Вместе**, в окне **Параметры гистограммы:** выделите режимы **Интервал между столбцами**, в окне **Интервалы:** выделите режим **Категории:** **6**, в окне **Подгонка:** выберите **Выкл** распределение (рис. 3.82) и нажмите ОК.

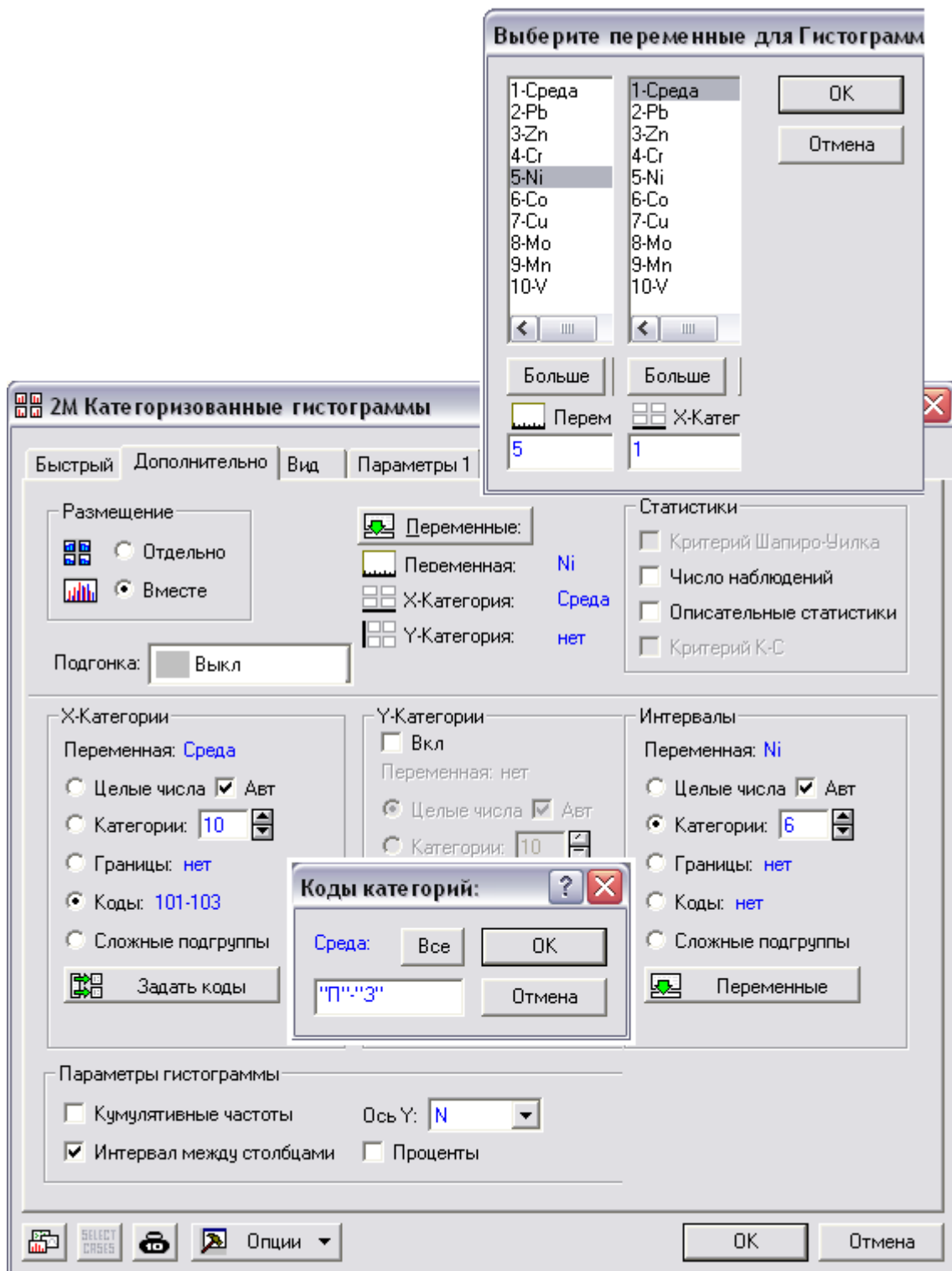


Рис. 3.82. Окно **2M Категоризованные гистограммы** для построения составной гистограммы Ni по средам

Следующая составная гистограмма Ni по средам П, С и З появится на экране (рис.3.83).

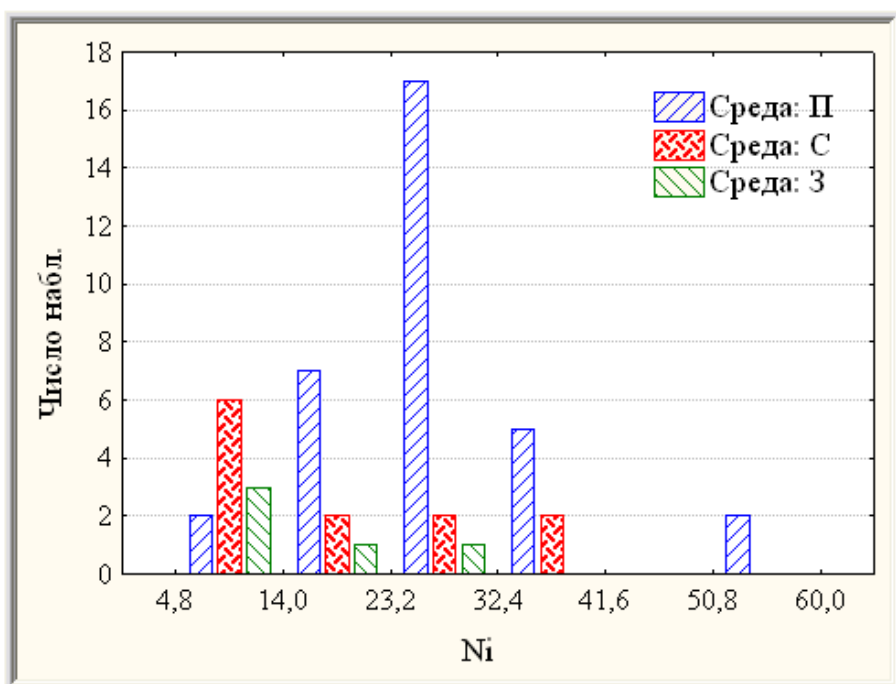


Рис. 3.83. Составная гистограмма распределения содержания микроэлемента Ni по средам П, С и З

Согласно табл.3.15, незначимо различаются распределения содержания Ni только по средам С и З.

Аналогичным образом строятся составные по средам П, С и З гистограммы распределений содержаний других химических элементов.

3.2.3. Корреляционный анализ содержаний микроэлементов в ассоциациях по трем средам.

Корреляция между результатами содержаний двух химических элементов может приблизительно характеризоваться прежде всего коэффициентом парной корреляции Пирсона r , отражающим степень линейной связи химических элементов, нормально распределенных. Последнее требование не всегда выполняется в случае малых выборок (см. рис. 3.75).

Для вычисления корреляционной матрицы ассоциаций необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль , затем нажмите

, выберите в появившемся окне (рис.3.84) и нажмите .

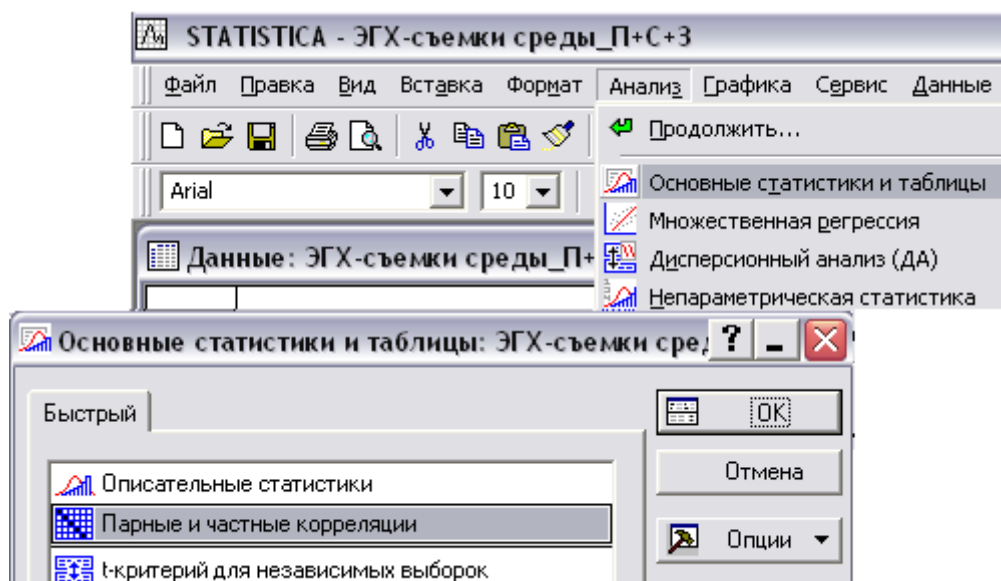


Рис. 3.84. Выбор **Парные и частные корреляции** в модуле **Анализ**

Шаг 2. В открывшемся окне **Парные и частные корреляции** нажмите левую кнопку **Квадратная матрица**, выберите все переменные среды П (почва) **раздельной по средам базы данных** и нажмите ОК (рис. 3.85). После этого нажмите кнопку **Матрица парных корреляций**.

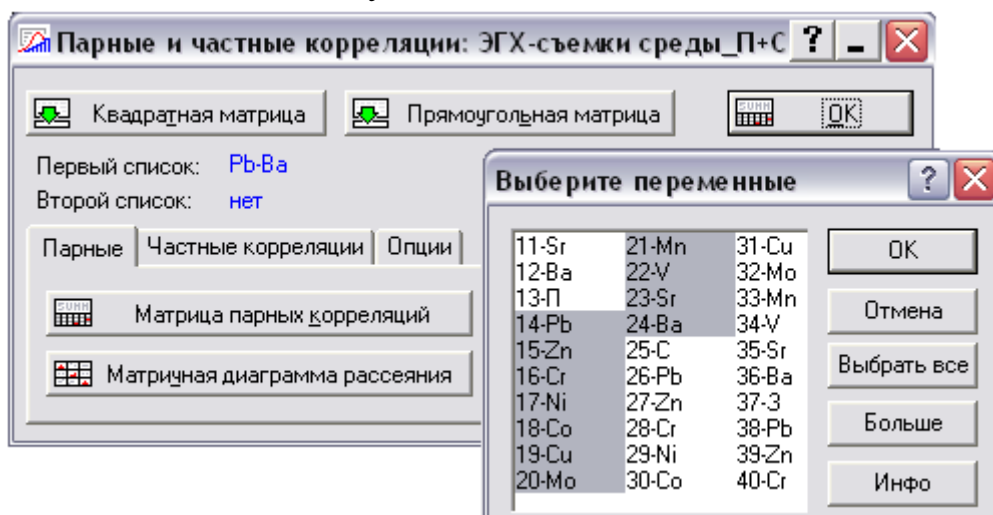


Рис. 3.85. Окно **Парные и частные корреляции**

Следующая таблица с результатами (значимые, то есть отличные от нуля, на уровне 0,05 коэффициенты парной корреляции Пирсона выделены красным цветом) появится на экране (табл.3.16).

Таблица 3.16

Корреляционная матрица ассоциации микроэлементов в почве (II) района г. Стрежевого.

(II)	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
Pb	1,00	0,73	-0,16	0,24	-0,04	0,30	-0,08	0,43	0,13	0,07	-0,05
Zn	0,73	1,00	-0,15	0,20	-0,01	0,44	-0,10	0,50	0,11	-0,04	0,01
Cr	-0,16	-0,15	1,00	-0,03	-0,15	0,04	0,13	-0,28	0,00	-0,06	-0,12
Ni	0,24	0,20	-0,03	1,00	0,78	0,40	0,65	0,62	0,57	0,31	0,49
Co	-0,04	-0,01	-0,15	0,78	1,00	0,19	0,54	0,55	0,66	0,53	0,63
Cu	0,30	0,44	0,04	0,40	0,19	1,00	0,32	0,62	0,25	0,29	0,07
Mo	-0,08	-0,10	0,13	0,65	0,54	0,32	1,00	0,28	0,38	0,26	0,27
Mn	0,43	0,50	-0,28	0,62	0,55	0,62	0,28	1,00	0,40	0,31	0,36
V	0,13	0,11	0,00	0,57	0,66	0,25	0,38	0,40	1,00	0,51	0,60
Sr	0,07	-0,04	-0,06	0,31	0,53	0,29	0,26	0,31	0,51	1,00	0,55
Ba	-0,05	0,01	-0,12	0,49	0,63	0,07	0,27	0,36	0,60	0,55	1,00

Аналогичным образом строятся корреляционные матрицы ассоциации микроэлементов в снеговом осадке (табл.3.17) и золе растений (табл.3.18) района г. Стрежевого.

Таблица 3.17

Корреляционная матрица ассоциации микроэлементов в снеговом осадке (C) района г. Стрежевого.

(C)	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
Pb	1,00	0,41	0,05	0,42	0,28	-0,00	0,35	0,13	0,63	0,19	0,69
Zn	0,41	1,00	0,65	0,90	0,70	0,58	0,51	0,88	0,44	0,09	0,52
Cr	0,05	0,65	1,00	0,48	0,25	0,62	0,71	0,72	0,06	-0,26	-0,04
Ni	0,42	0,90	0,48	1,00	0,78	0,54	0,46	0,81	0,53	0,30	0,68
Co	0,28	0,70	0,25	0,78	1,00	0,39	0,46	0,45	0,31	0,66	0,65
Cu	-0,00	0,58	0,62	0,54	0,39	1,00	0,63	0,67	0,04	-0,12	0,22
Mo	0,35	0,51	0,71	0,46	0,46	0,63	1,00	0,44	0,23	-0,01	0,34
Mn	0,13	0,88	0,72	0,81	0,45	0,67	0,44	1,00	0,29	-0,19	0,23
V	0,63	0,44	0,06	0,53	0,31	0,04	0,23	0,29	1,00	0,13	0,77
Sr	0,19	0,09	-0,26	0,30	0,66	-0,12	-0,01	-0,19	0,13	1,00	0,52
Ba	0,69	0,52	-0,04	0,68	0,65	0,22	0,34	0,23	0,77	0,52	1,00

Таблица 3.18

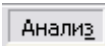
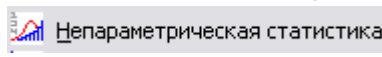

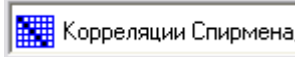
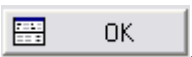
Корреляционная матрица ассоциации микроэлементов в золе (3) расте-
ний района г. Стрежевого.

(3)	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
Pb	1,00	0,33	0,41	-0,24	-0,55	-0,24	-0,94	0,32	0,49	-0,86	-0,51
Zn	0,33	1,00	-0,22	0,17	-0,00	0,26	-0,17	0,33	-0,00	-0,67	-0,76
Cr	0,41	-0,22	1,00	0,58	0,27	0,50	-0,33	0,64	0,98	-0,22	-0,07
Ni	-0,24	0,17	0,58	1,00	0,85	0,95	0,37	0,70	0,63	0,10	-0,11
Co	-0,55	-0,00	0,27	0,85	1,00	0,69	0,52	0,61	0,28	0,52	0,34
Cu	-0,24	0,26	0,50	0,95	0,69	1,00	0,46	0,50	0,57	-0,03	-0,34
Mo	-0,94	-0,17	-0,33	0,37	0,52	0,46	1,00	-0,33	-0,37	0,67	0,22
Mn	0,32	0,33	0,64	0,70	0,61	0,50	-0,33	1,00	0,73	-0,22	-0,07
V	0,49	-0,00	0,98	0,63	0,28	0,57	-0,37	0,73	1,00	-0,37	-0,24
Sr	-0,86	-0,67	-0,22	0,10	0,52	-0,03	0,67	-0,22	-0,37	1,00	0,87
Ba	-0,51	-0,76	-0,07	-0,11	0,34	-0,34	0,22	-0,07	-0,24	0,87	1,00

Жирным шрифтом выделены наименьший из значимых коэффициентов парной корреляции Пирсона и наиболее значимый коэффициент корреляции.

Как следует из табл.3.16 – 18, наиболее значимые корреляционные связи в ассоциациях наблюдаются между Ni и Co в почве, между Ni и Zn в снеговом осадке, между V и Cr в золе растений. Учитывая отсутствие корреляционной связи между V и Cr в почве ($r = 0,00$ в табл.3.16) и в снеговом осадке ($r = 0,06$ по табл. 3.17), рассмотрим вместо нее следующую по значимости пару в золе Ni и Cu. Рассмотрим эти наиболее значимые корреляционные связи в других средах.

В случае выборок малого объема корреляция между результатами содержания двух микроэлементов более корректно может характеризоваться ранговым коэффициентом корреляции Спирмена R . Для определения коэффициента корреляции Спирмена R необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль  , в стартовой панели которого нажмите  (рис.3.86). В появившемся окне  выберите  (рис. 3.86) и нажмите .

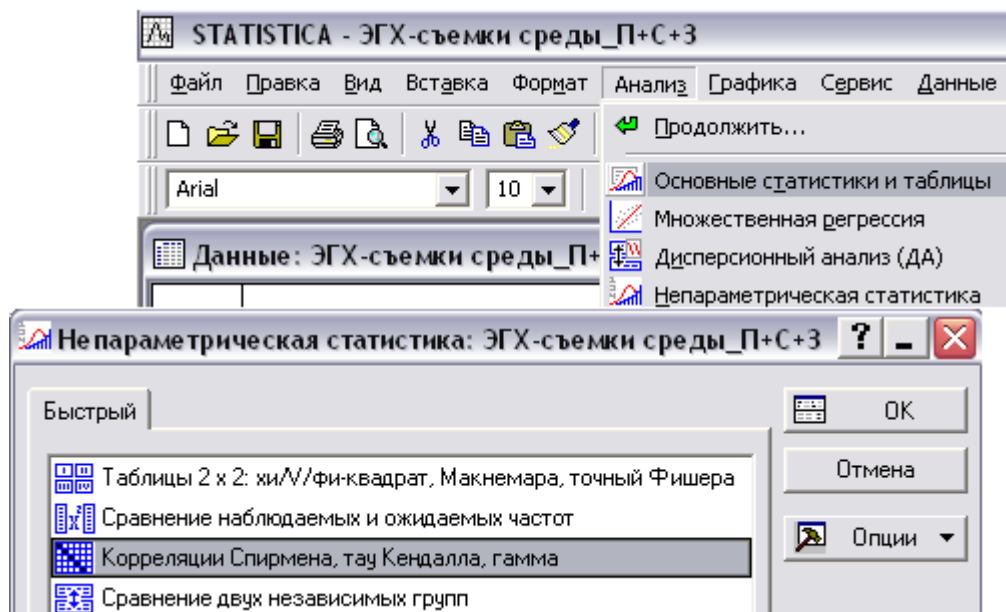


Рис. 3.86. Выбор модуле **Анализ**

Шаг 2. В открывшемся окне **Ранговые корреляции** выберите режим вычисления **Вычислить: Подробный отчет**, далее нажмите кнопку **Переменные** и выберите два набора переменных базы данных среды П (например, «Ni» – в окне **Первый список**, а Co, Zn, Cu – в окне **Второй список**) и нажмите **OK** (рис.3.87). Затем нажмите кнопку **Спирмена R**.

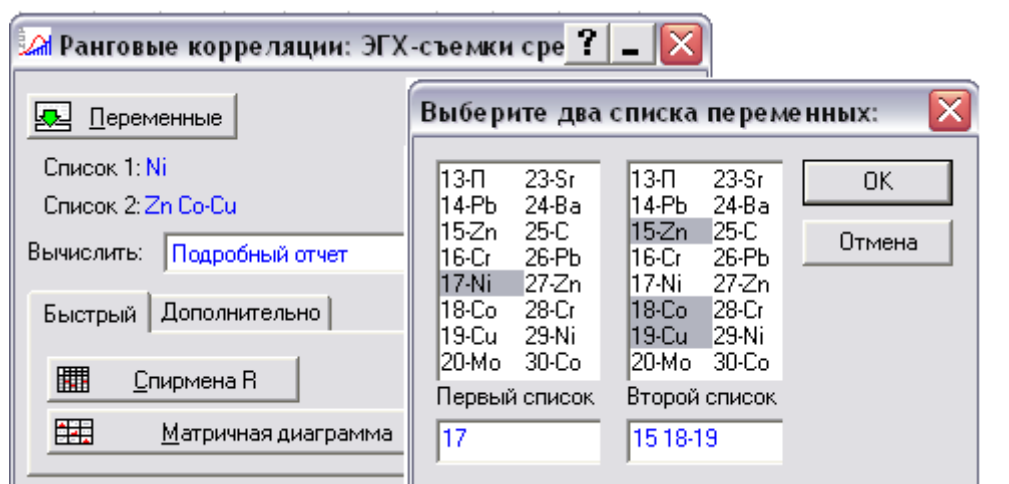


Рис. 3.87. Окно **Ранговые корреляции** в режиме **Подробный отчет**

На экране появится заказанный фрагмент корреляционной матрицы Спирмена R в следующем формате (рис.3.88):

Пара перем.	Ранговые корреляции Спирмена (ЭГХ-съемки среды ПД попарно удалены) Отмеченные корреляции значимы на уровне $p < ,050$			
	Число набл.	Спирмена R	t(N-2)	p-уров.
Ni & Zn	33	0,388	2,342	0,02581
Ni & Co	33	0,792	7,227	0,00000
Ni & Cu	33	0,603	4,206	0,00021

Рис. 3.88. Фрагмент корреляционной матрицы Спирмена R

В данном случае наряду со значениями коэффициентов парной корреляции Спирмена R вычислены также соответствующие критерии $t(N-2)$ проверки нулевой гипотезы $H_0: R = 0$ (аналогично (1.34)) и их p -уровни значимости (отличия от нуля).

Все рассмотренные выше случаи по средам можно свести в табл.3.19.

Таблица 3.19

Коэффициенты корреляции Спирмена для наиболее значимых корреляционных связей по средам района г. Стржевого.

Пары элементов	П			С			З		
	n	R	p-level	n	R	p-level	n	R	p-level
Ni & Co	33	0,792	0,0000	12	0,578	0,049	5	0,872	0,054
Ni & Zn	33	0,388	0,026	12	0,738	0,0062	5	0,289	0,64
Ni & Cu	33	0,603	0,0002	12	0,622	0,031	5	0,949	0,0138

Далее можно оценить значимости различия между двумя коэффициентами корреляции одной и той же пары микроэлементов в разных средах. Для этого необходимо:

Шаг 1. Запустите в головном меню модуль **Анализ**, выберите

 Основные статистики и таблицы, затем выберите  Другие критерии значимости (рис.

3.89) и нажмите .

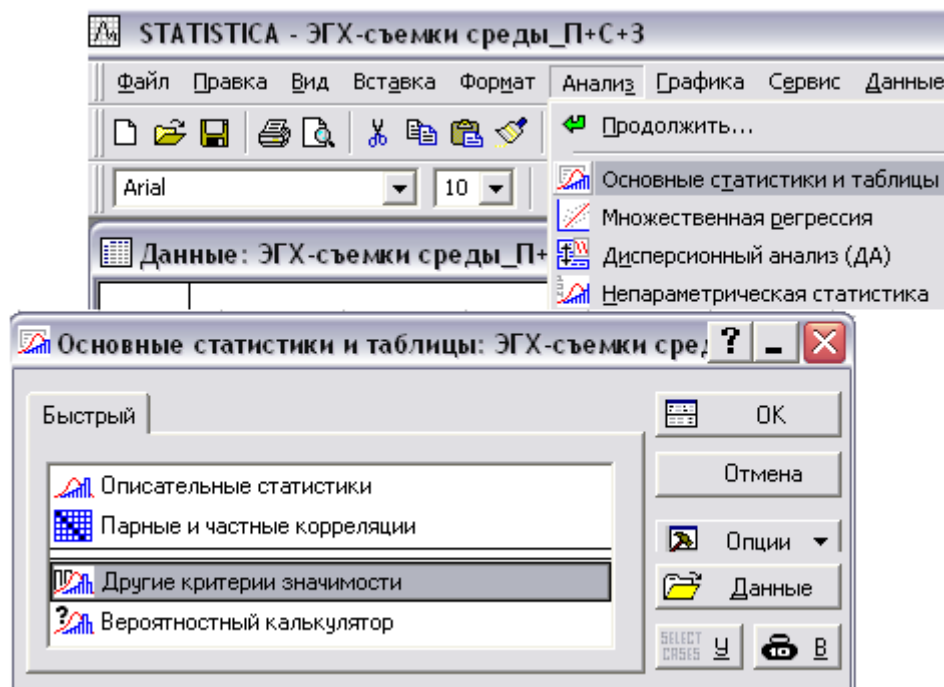


Рис. 3.89. Выбор модуля **Анализ** - **Основные статистики и таблицы**

Шаг 2. В окне **Другие критерии значимости** введите значения сравниваемых значений коэффициентов корреляции (напр., для пары N_1 & S_0 соответственно $r_1 = 0,79$ и $r_2 = 0,58$), объемов выборок ($N_1 = 33$ и $N_2 = 12$), выделите точкой двухсторонний режим и нажмите **Вычислить** (рис. 3.90). Здесь же в окне появится значение $p = 0,2882$, то есть данные $r_1 = 0,79$ и $r_2 = 0,58$ при $N_1 = 33$ и $N_2 = 12$ различаются незначимо.

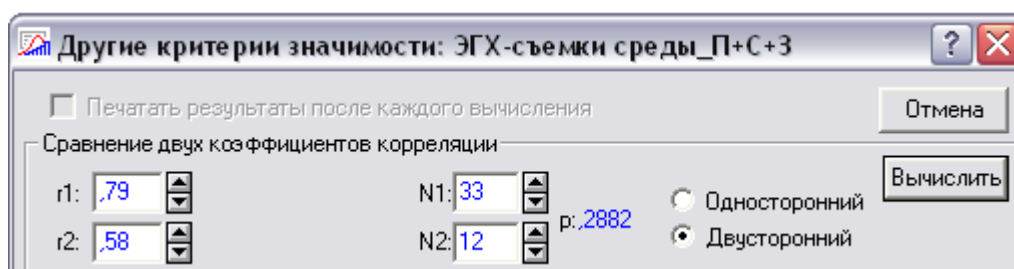


Рис. 3.90. Оценка расхождения между двумя коэффициентами корреляции Пирсона

Перебирая таким образом все наиболее значимые пары (см. табл.3.19), построим таблицу значимости различия между двумя коэффициентами корреляции одной и той же пары микроэлементов в разных средах (табл.3.20).

Таблица 3.20

Оценка значимости различия между двумя коэффициентами корреляции одной и той же пары микроэлементов в разных средах.

Элементы	П - С	П - З	С - З
Ni & Co	0,29	0,72	0,41
Ni & Zn	0,16	0,88	0,42
Ni & Cu	0,93	0,13	0,18

Таким образом, различия наиболее значимых корреляционных связей являются незначимыми по средам. В значительной степени это обстоятельство объясняется малыми объемами выборок. Например, если для пары Ni & Cu в случае С – З: $r_1 = 0,62$ и $r_2 = 0,95$ при $N_1 = 12$ и $N_2 = 5$ значение $p = 0,18$ дает различия незначимые, то в случае С – З: $r_1 = 0,62$ и $r_2 = 0,95$ при $N_1 = 12$ и $N_2 = 55$ (отличие только в N_2) уже значение $p = 0,003$ дает различия статистически значимые.

3.2.4. Регрессионный анализ содержаний микроэлементов в ассоциациях по трем средам.

Построим составные диаграммы рассеяния и уравнения регрессий для наиболее значимых корреляционных связей в ассоциациях:

Шаг 1. Запустите в головном меню модуль **Графика**, затем выберите **Категоризованные графики** и далее **Диаграммы рассеяния...** (рис. 3.91).

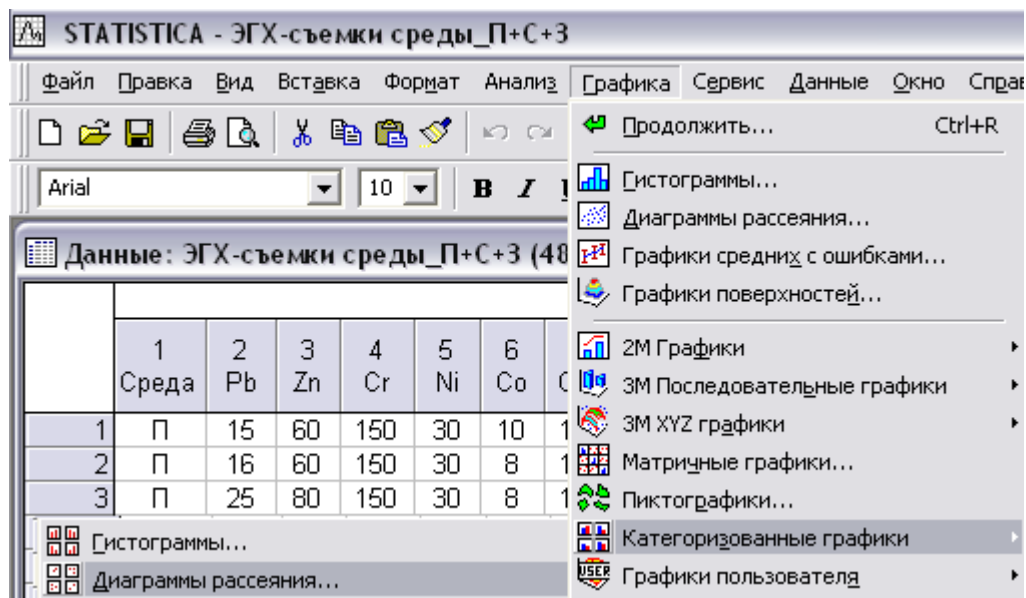


Рис. 3.91. Выбор **Диаграммы рассеяния...** в модуле **Графика**.

Шаг 2. В появившемся окне щелкните **Дополнительно** (рис. 3.92). В открывшемся окне нажмите кнопку **Переменные:** и выберите переменные категориальной базы данных (например, переменная «Ni» – в окне **X:**, переменная «Cu» – в окне **Y:** и переменная «Среда» – в окне **X-Категория:**) и нажмите **OK**. Далее в окне **X-Категории** выделите режим **Коды:**, нажмите кнопку **Задать коды**, выберите **Все** в окне **Коды категорий:** и нажмите **OK**. После этого в окне **Статистики** выделите режим **Корреляция и r-уров.**, в окне **Тип подгонки:** – режим **Линейная**, а в окне **Размещение** – режим **Вместе** (рис. 3. 93) или **Отдельно** (рис. 3. 94) и нажмите **OK**.

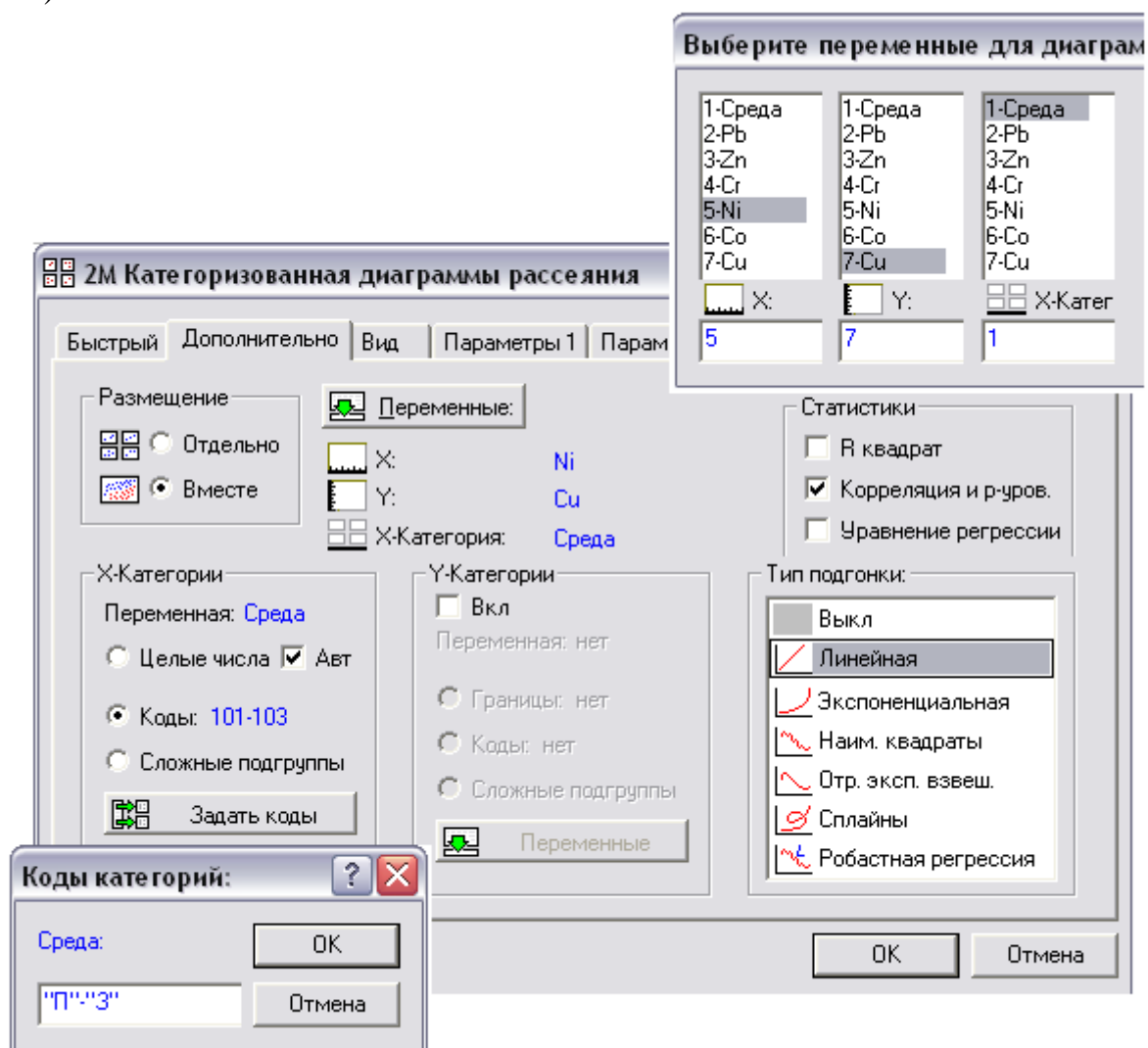


Рис. 3.92. Окно **2M Категоризованная диаграммы рассеяния**

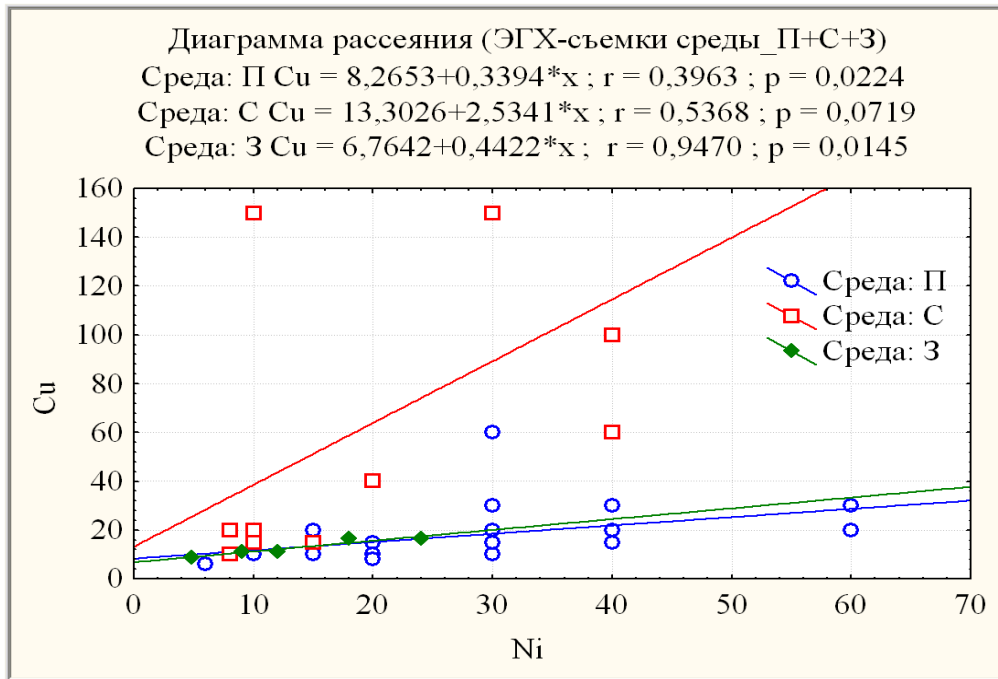


Рис.3.93. Составная диаграмма рассеяния Ni и Cu по средам П, С и З.

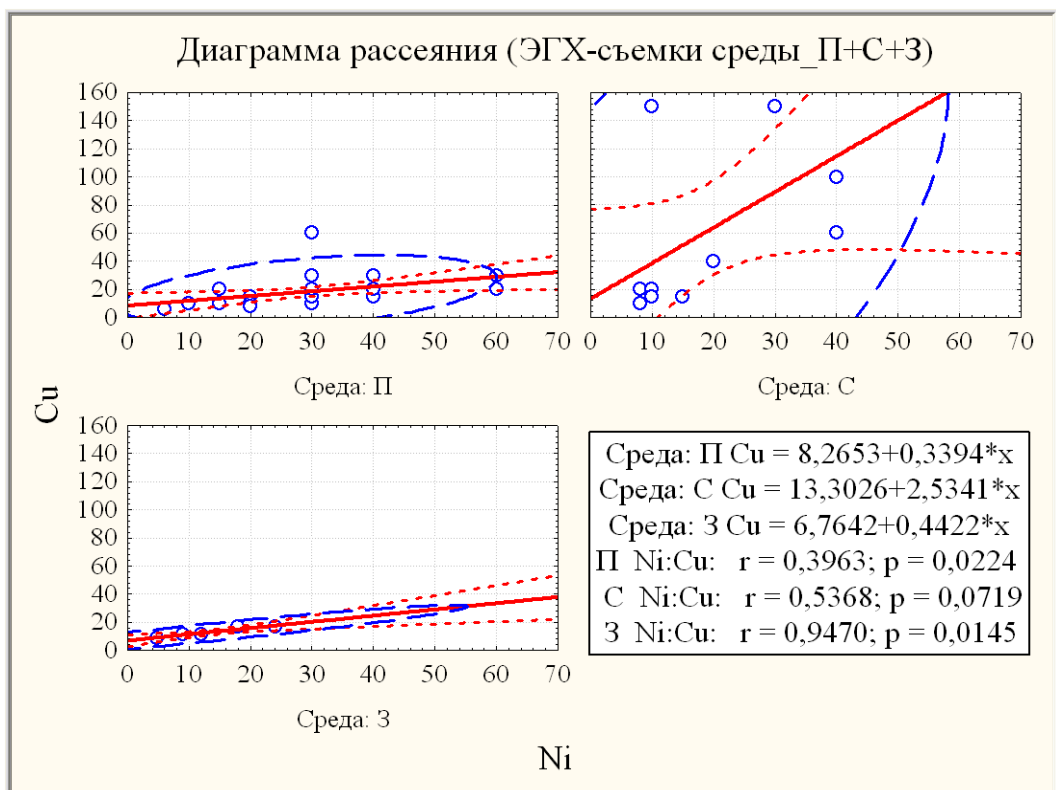







Рис.3.94. Раздельные диаграммы рассеяния Ni и Cu по средам П,С,З.

В последнем случае к прямой регрессии добавлены 95% эллипс рассеяния и 95% доверительный интервал.

3.2.5. Кластерный анализ содержаний микроэлементов в ассоциациях по трем средам.

Рассмотрим применение кластерного анализа для сопоставления результатов исследований по средам и получения достоверных отличий посредством сопоставления ассоциаций химических элементов в дендрограммах разных выборок по средам. Результаты кластер – анализа изображаются в виде дендрограммы, в которой по оси абсцисс располагаются символические обозначения химических элементов, а по оси ординат значение $1 - r$, соответствующее каждому иерархическому уровню группирования.

Для оценки сходства результатов, полученных в разных средах, на основе кластерного анализа (метода иерархической группировки) необходимо выполнить следующие действия:

Шаг 1. Выберите  Многомерный разведочный анализ в стартовой панели модуля  и далее щелкните мышкой на  (рис.3.95). В появившемся окне выделите разновидность  и щелкните  (рис. 3.95).

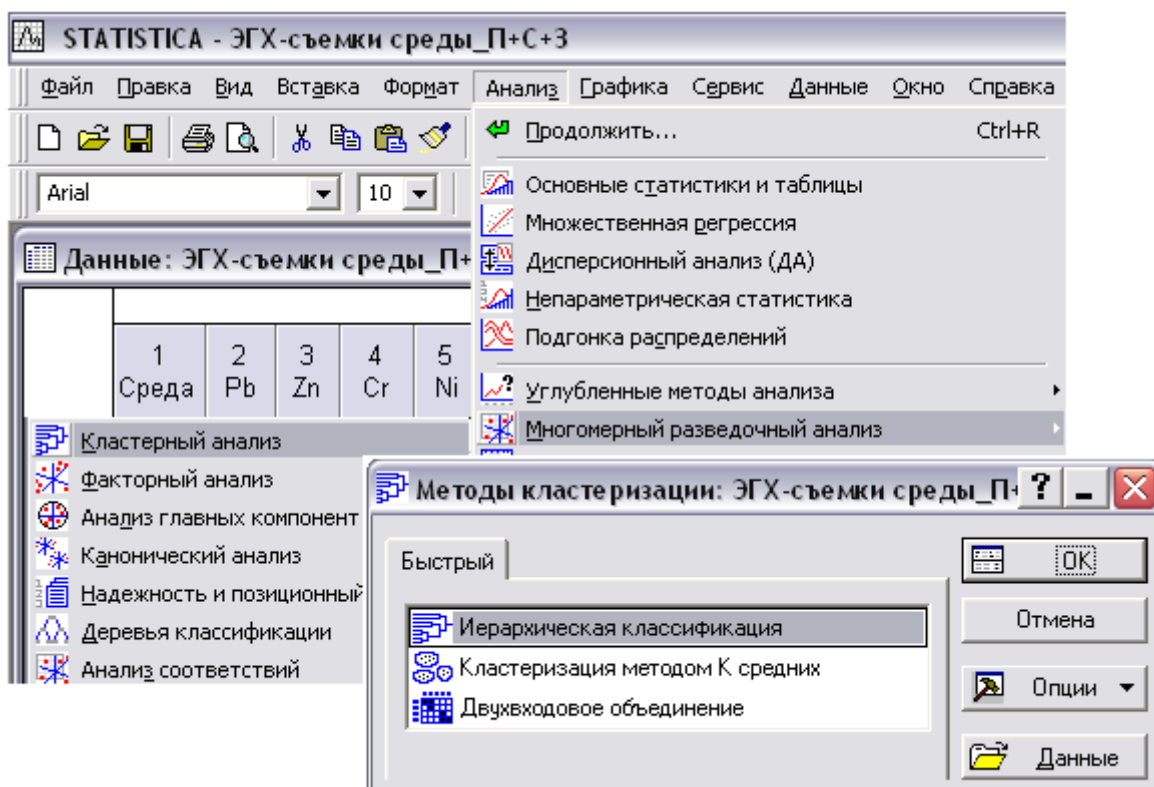


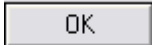

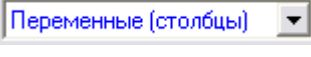







Рис. 3.95. Выбор  Методы кластеризации: в модуле 

Шаг 2. Нажмите кнопку  Переменные: в открывшемся окне  Кластерный анализ: нерархическая классификация: и выберите переменные раздельной базы данных П (рис. 3.96) и нажмите .

Шаг 3. Далее в окне  Объекты: выделите , в окне  Мера близости: – , а в окне  Правило объединения: –  и нажмите  (рис. 3.96).

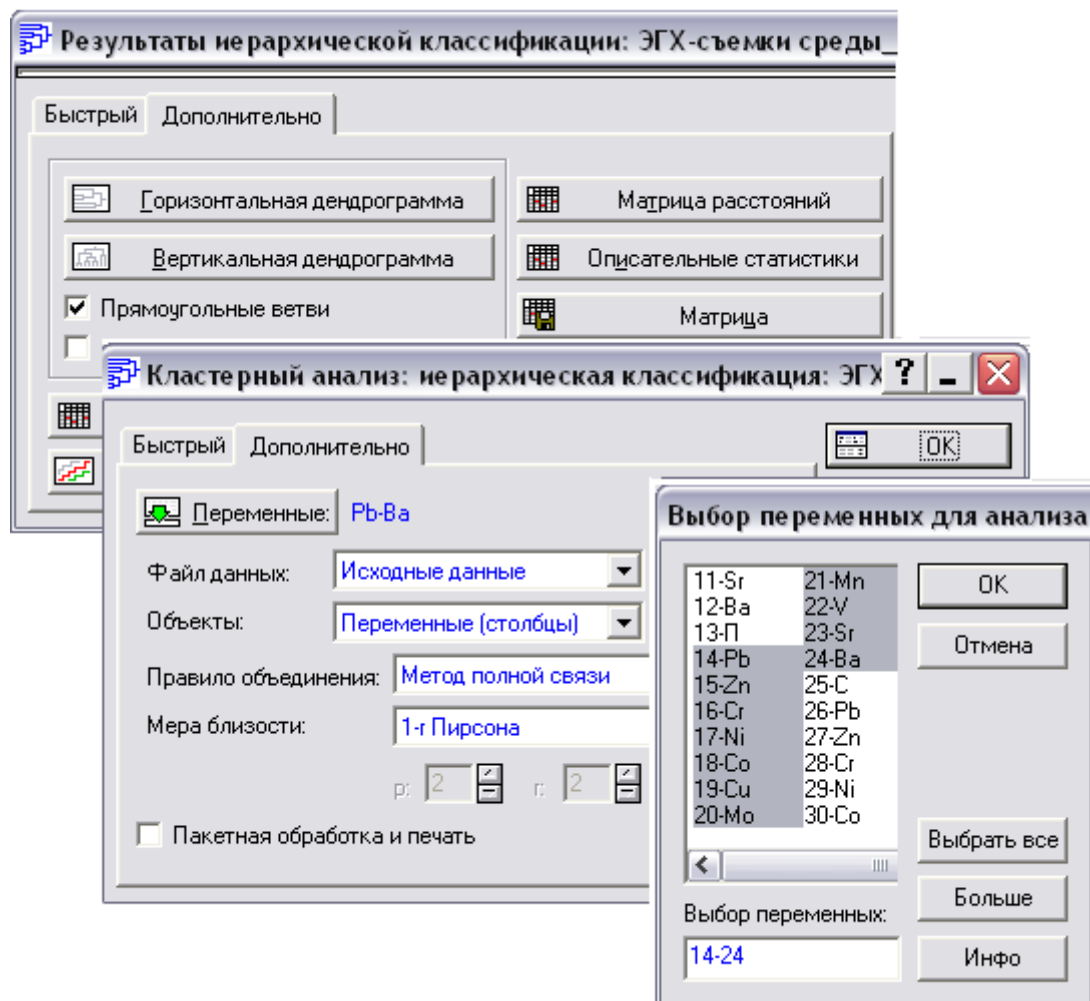




Рис. 3.96. Окно  Кластерный анализ: нерархическая классификация:

Шаг 4. После этого щелкните  Вертикальная дендрограмма в открывшемся окне  Результаты нерархической классификации: (рис. 3.96).

Следующая дендрограмма по среде П появится на экране (рис.3.97). Аналогично – по С (рис.3.98) и – по З (рис.3.99).

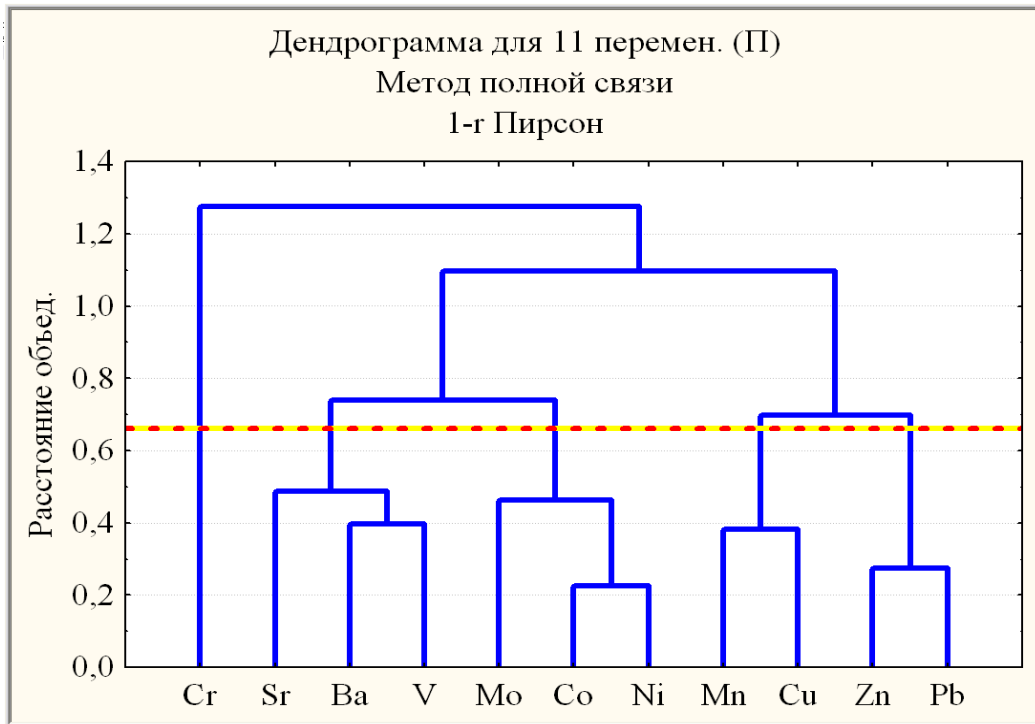


Рис. 3.97. Дендрограмма корреляционной матрицы геохимического спектра микроэлементов в почве ($1 - r_{0,05} = 0,66; n = 33$)

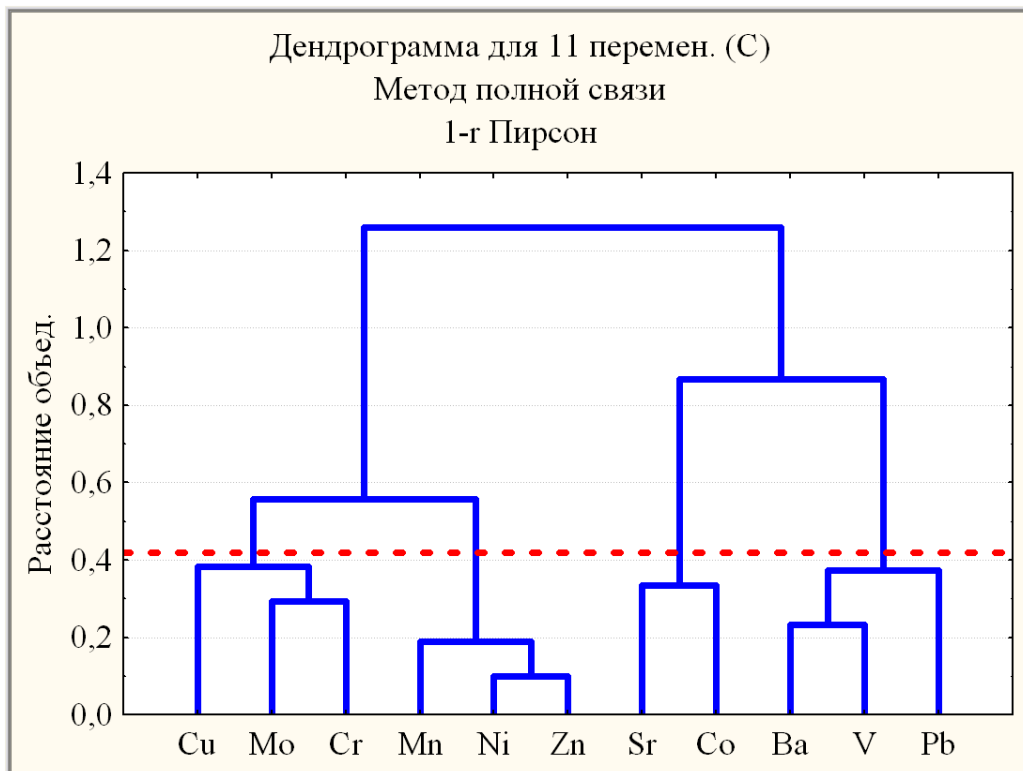


Рис. 3.98. Дендрограмма корреляционной матрицы геохимического спектра микроэлементов в снеговом осадке ($1 - r_{0,05} = 0,42; n = 12$)

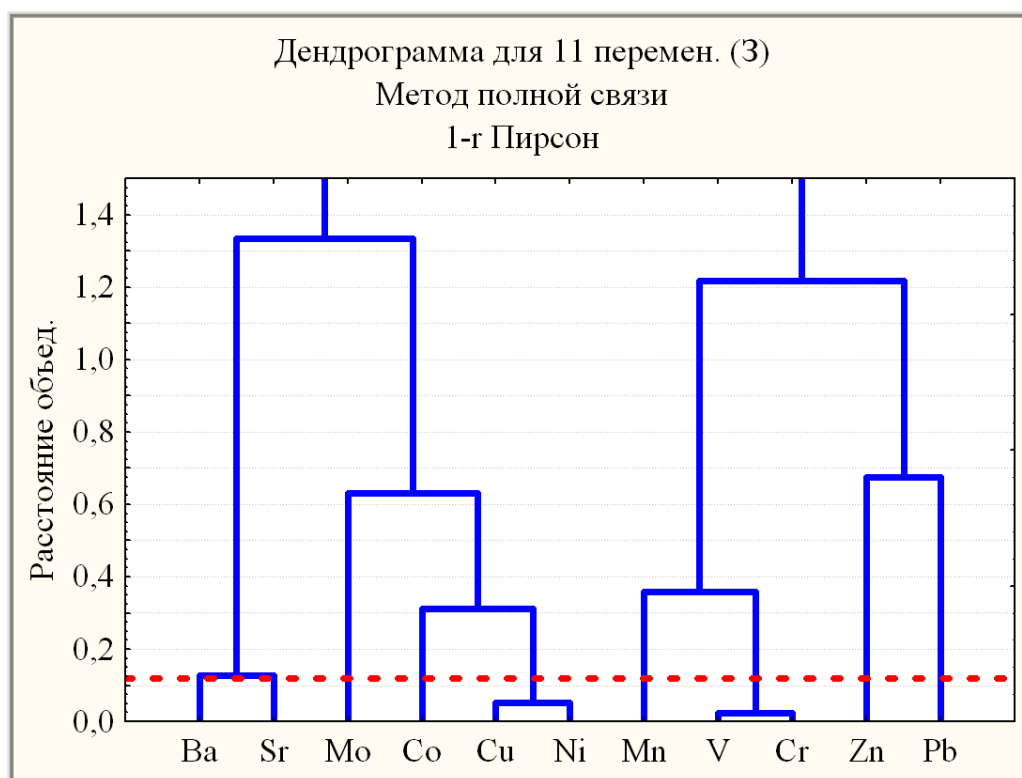


Рис. 3.99. Дендрограмма корреляционной матрицы геохимического спектра микроэлементов в золе растений ($1 - r_{0,05} = 0,12$; $n = 5$).

Здесь дополнительно добавлена пунктирная прямая критического уровня (в почве $1 - r_{0,05} = 0,66$, в снеговом осадке $1 - r_{0,05} = 0,42$ и в золе растений $1 - r_{0,05} = 0,12$).

Для изображения горизонтальной пунктирной прямой критического уровня необходимо выполнить последовательность следующих действий:

Шаг 1. Дважды щелкнуть мышкой по дендрограмме для вхождения в модуль **Все параметры графика** (рис. 3.100).

Шаг 2. Активировать **Пользовательская функция** и нажать **Новая функция**.

Шаг 3. Заказать ряд параметров в окнах справа: в окне **Тип:** - тип **Параметр. кривая**, в окне **Браузер функций** - задание параметрической функции $\{ Y(t) = t, X(t) = 0,66 \text{ в почве} \}$, в окне **Линия** задать свойства линии с помощью кнопки **Шаблон...** - выбрать **Толщина линии:**, **Шаблон линии:** (пунктирный), **Передний фон:** (красный цвет) и **Задний фон:** (желтый цвет) линии, в окне **Диапазон изменения параметра t** - значения 0 – 12, где число 12 - число химических элементов плюс 1 (рис. 3.100)

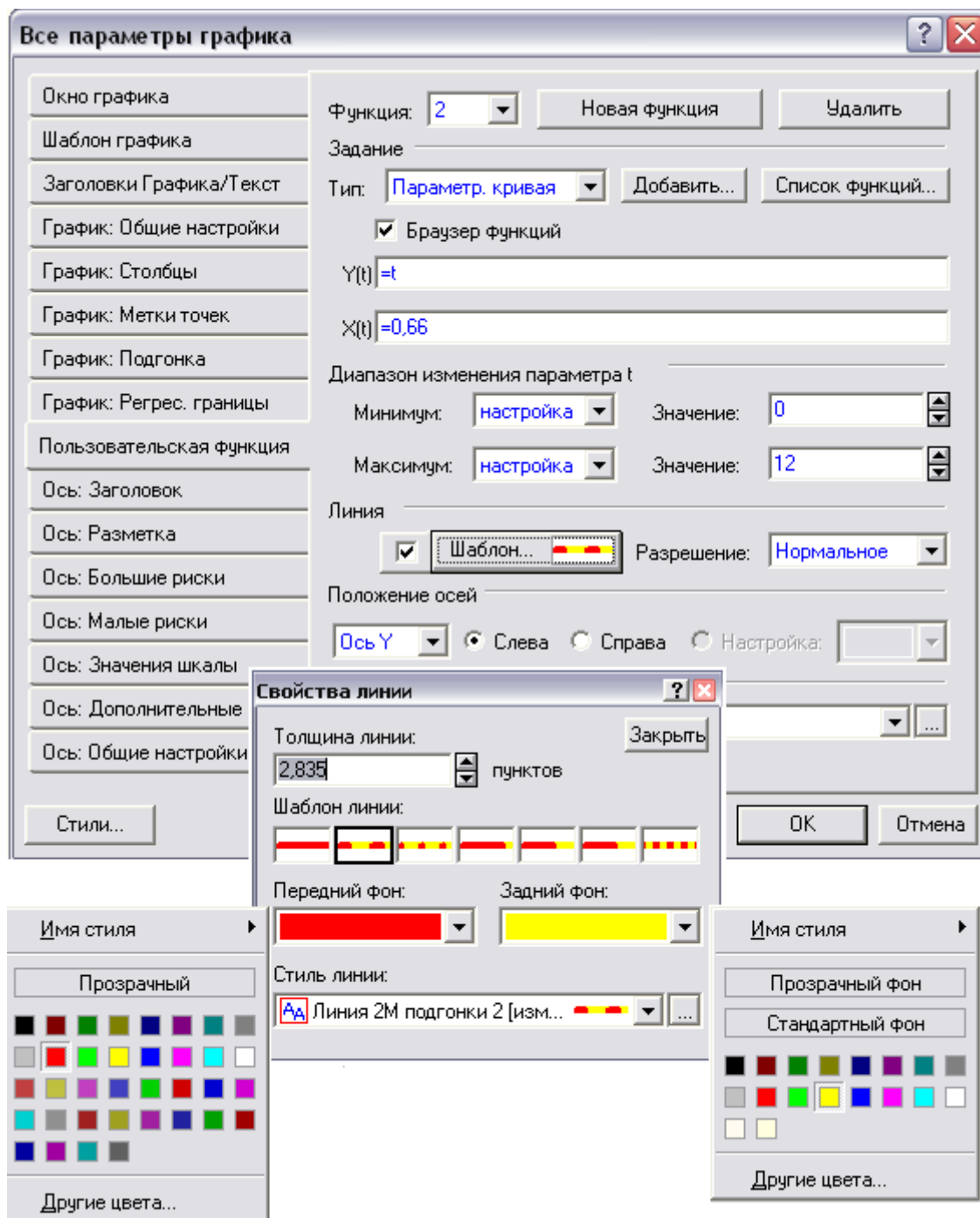


Рис. 3.100. Режим Пользовательская функция в модуле Все параметры графика

Как следует из рис.3.97-99, значимые (на уровне значимости $p < 0,05$) геохимические спектры микроэлементов в почве, в снеговом осадке в золе растений существенно различаются:

II: {V–Ba–Sr; Ni–Co–Mo; Mn–Cu; Zn–Pb} – 4-е составных кластера и 1 моно кластер {Cr},

C: { V–Ba– Pb; Cr– Mo– Cu; Zn– Ni – Mn; Co– Sr; } – 4-е составных кластера,

З: {Cu–Ni; V–Cr} – 2-а составных кластера, остальные микроэлементы образуют моно кластеры.

Наиболее явно (по количеству значимых кластеров) наблюдается отличие среды З от остальных, но даже в случае сред П и С, имеющих по 4-е значимых составных кластера, наблюдается очень скромное частичное пересечение одной пары кластеров по {V–Ba}.

Наличие корреляционной зависимости микроэлементов допускает использование факторного анализа (п. 1.2.5), основанного на оценках корреляций (факторных нагрузок) между исходными микроэлементами и факторами, образованными группами корреляционно связанных исходных микроэлементов.

3.2.6. Факторный анализ содержаний микроэлементов в снеговом осадке.

Главными целями факторного анализа являются сокращение числа рассматриваемых микроэлементов и определение структуры взаимосвязей между микроэлементами, т.е. классификация микроэлементов.

С использованием пакета STATISTICA 6.1 данная задача может быть решена в следующей последовательности.

Шаг 1. Запустите **Анализ**, нажмите **Многомерный разведочный анализ**, где далее выберите **Факторный анализ** (рис. 3.101).

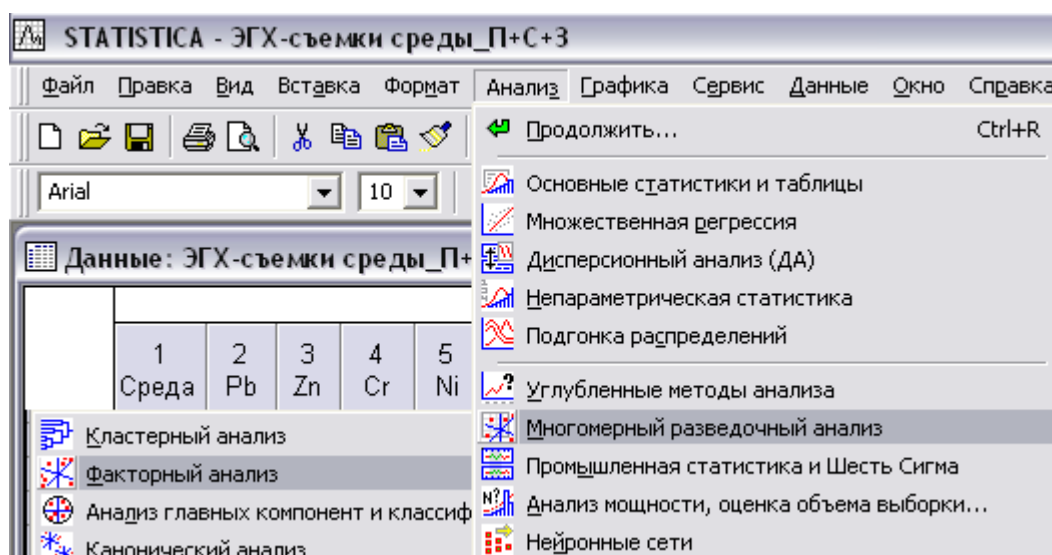
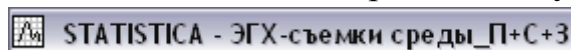


Рис. 3.101. Выбор **Факторный анализ** в стартовом модуле **Анализ** системы



Шаг 2. Выберите  Переменные: (рис. 3.102) отдельной базы данных среды С в открывшемся окне  Факторный анализ: и нажмите , а затем . В результате появится рис. 3.103.

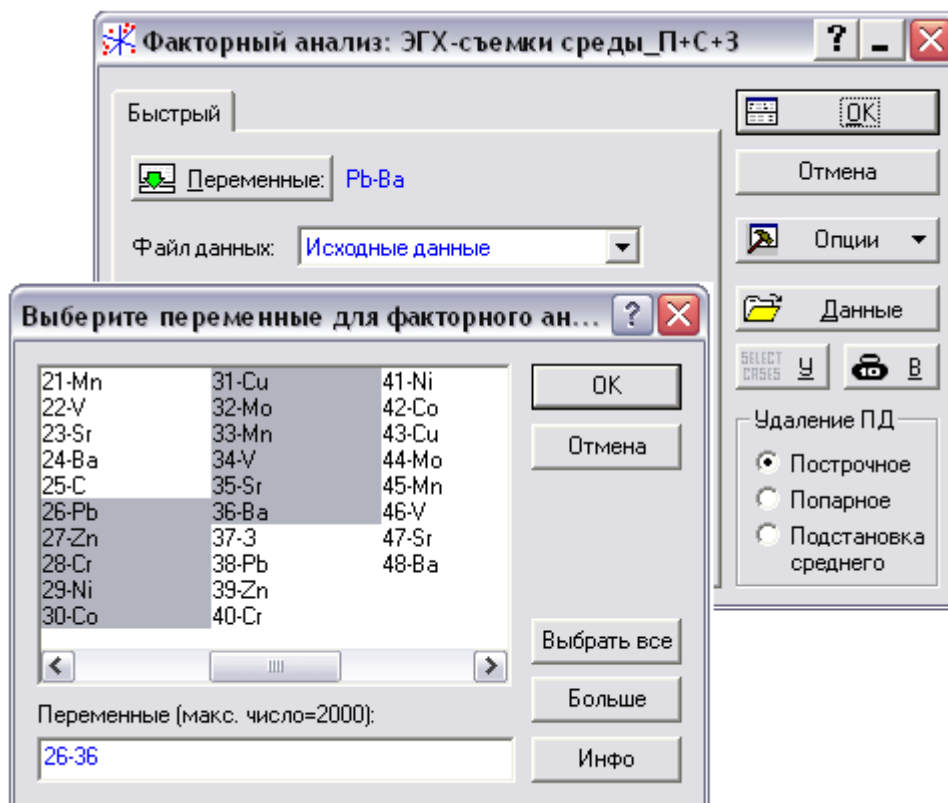


Рис. 3.102. Выбор переменных среды С для факторного анализа

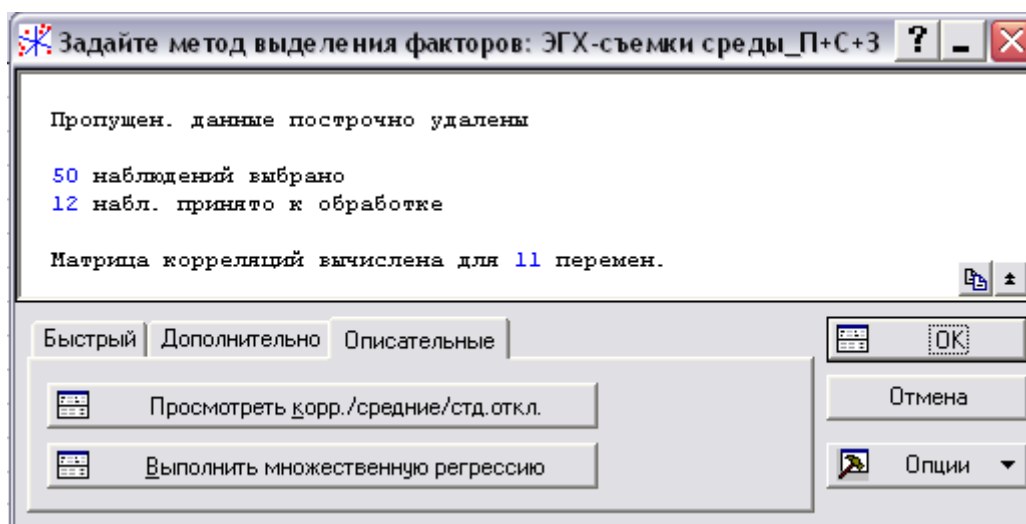

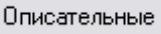


Рис. 3.103. Выбор режима  Описательные для просмотра описательной статистики

Шаг 3. Нажмите кнопку  Просмотреть корр./средние/стд.откл. в режиме  Описательные (рис. 3.103). В результате появится рис. 3.104.

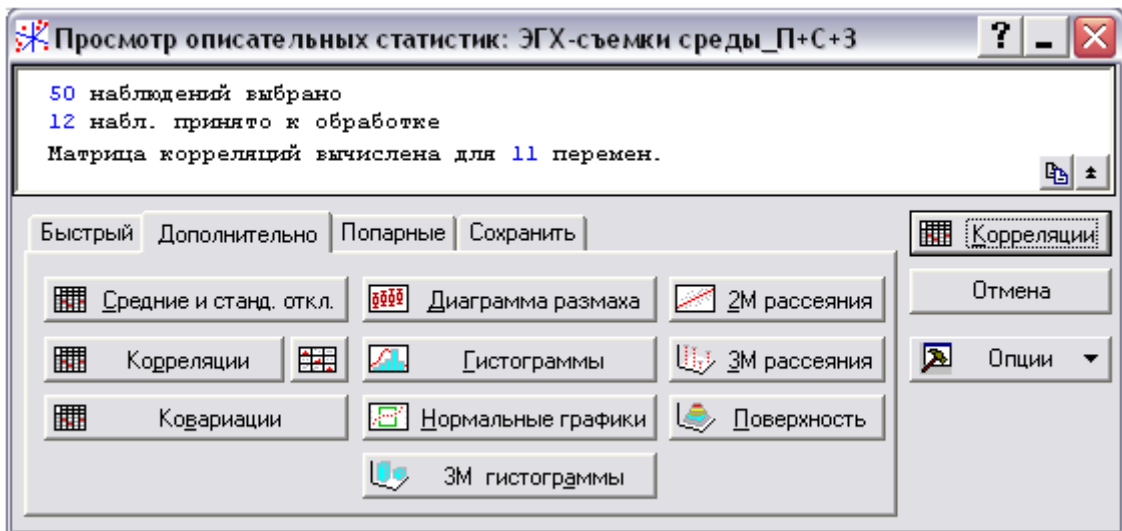
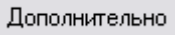
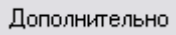

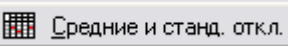
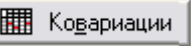
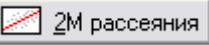
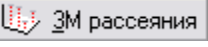


Рис.3.104. Окно для описательной статистики в режиме  Дополнительно

Можно посмотреть (рис.3.104) в режиме  Дополнительно уже знакомую корреляционную матрицу  Корреляции (см. табл.3.17) и прочую описательную статистику выборки ( Средние и станд. откл. и  Ковариации), построить различные графики, например,  2М рассеяния Ni – Mn (рис. 3.105) или  3М рассеяния Ni – Zn – Mn (рис. 3.106).

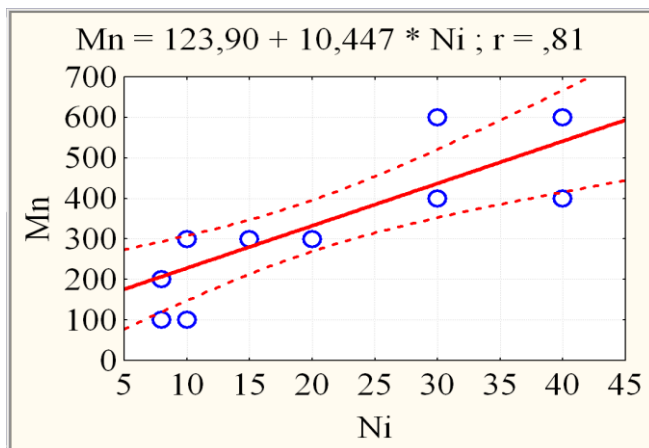


Рис.3.105. 2М Диаграмма рассеяния Ni – Mn

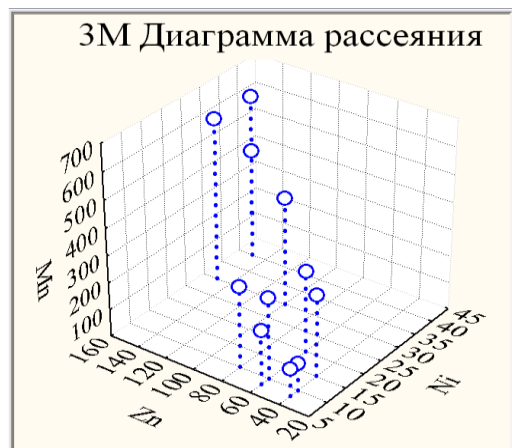
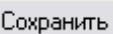


Рис.3.106. 3М Диаграмма рассеяния Ni – Zn – Mn

Описательную статистику можно посмотреть также в режиме

 Сохранить (рис. 3.107)

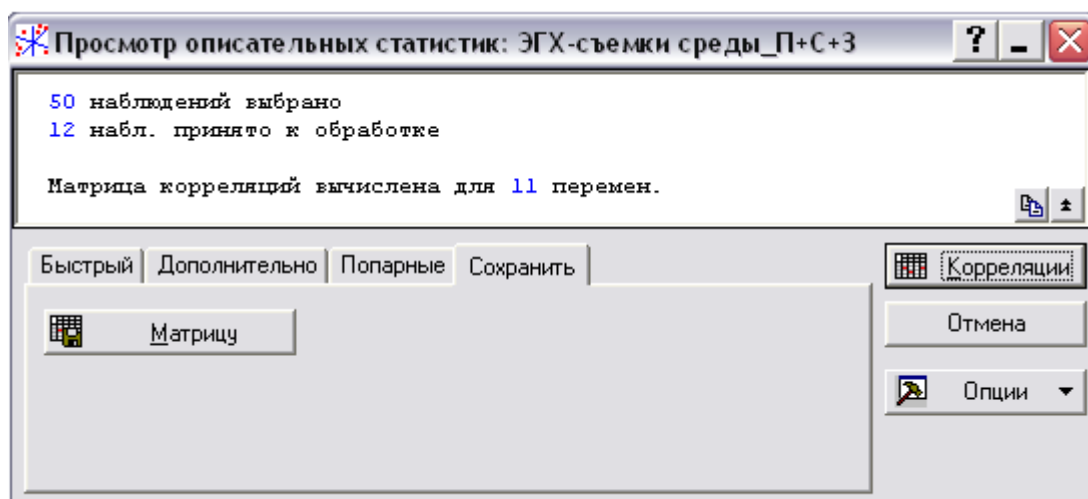


Рис.3.107. Окно для просмотра описательной статистики в режиме

Сохранить

Шаг 4. В режиме **Сохранить** (рис. 3.107) нажмите **Матрицу** для вывода на экран корреляционной матрицы, расширенной строками средних и стандартных отклонений (рис. 3.108).

		ЭГХ-съемки среды_П+С+3										
		1	2	3	4	5	6	7	8	9	10	11
		Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
Pb		1,00	0,41	0,05	0,42	0,28	-0,00	0,35	0,13	0,63	0,19	0,69
Zn		0,41	1,00	0,65	0,90	0,70	0,58	0,51	0,88	0,44	0,09	0,52
Cr		0,05	0,65	1,00	0,48	0,25	0,62	0,71	0,72	0,06	-0,26	-0,04
Ni		0,42	0,90	0,48	1,00	0,78	0,54	0,46	0,81	0,53	0,30	0,68
Co		0,28	0,70	0,25	0,78	1,00	0,39	0,46	0,45	0,31	0,66	0,65
Cu		-0,00	0,58	0,62	0,54	0,39	1,00	0,63	0,67	0,04	-0,12	0,22
Mo		0,35	0,51	0,71	0,46	0,46	0,63	1,00	0,44	0,23	-0,01	0,34
Mn		0,13	0,88	0,72	0,81	0,45	0,67	0,44	1,00	0,29	-0,19	0,23
V		0,63	0,44	0,06	0,53	0,31	0,04	0,23	0,29	1,00	0,13	0,77
Sr		0,19	0,09	-0,26	0,30	0,66	-0,12	-0,01	-0,19	0,13	1,00	0,52
Ba		0,69	0,52	-0,04	0,68	0,65	0,22	0,34	0,23	0,77	0,52	1,00
Средние		99,17	82,50	69,17	19,25	5,17	62,08	3,58	325,00	32,50	258,33	391,67
Стд. откл		43,16	44,13	29,99	12,45	2,08	58,79	1,31	160,26	23,31	51,49	156,43
Кол-во N		12										

Рис. 3.108. Наблюдаемая корреляционная матрица, расширенная строками средних и стандартных отклонений выборки С

Шаг 5. Для проведения факторного анализа выборки вернитесь в окно **Задайте метод выделения факторов** (рис. 3.103) с помощью кнопки **Отмена** и выберите режим **Дополнительно** (рис. 3.109), в котором в поле справа можно указать Максимальное число факторов, которые будут выделены системой, например, в начале исследования 11, и Минимальное собственное значение, например 0,00 (рис. 3.109).

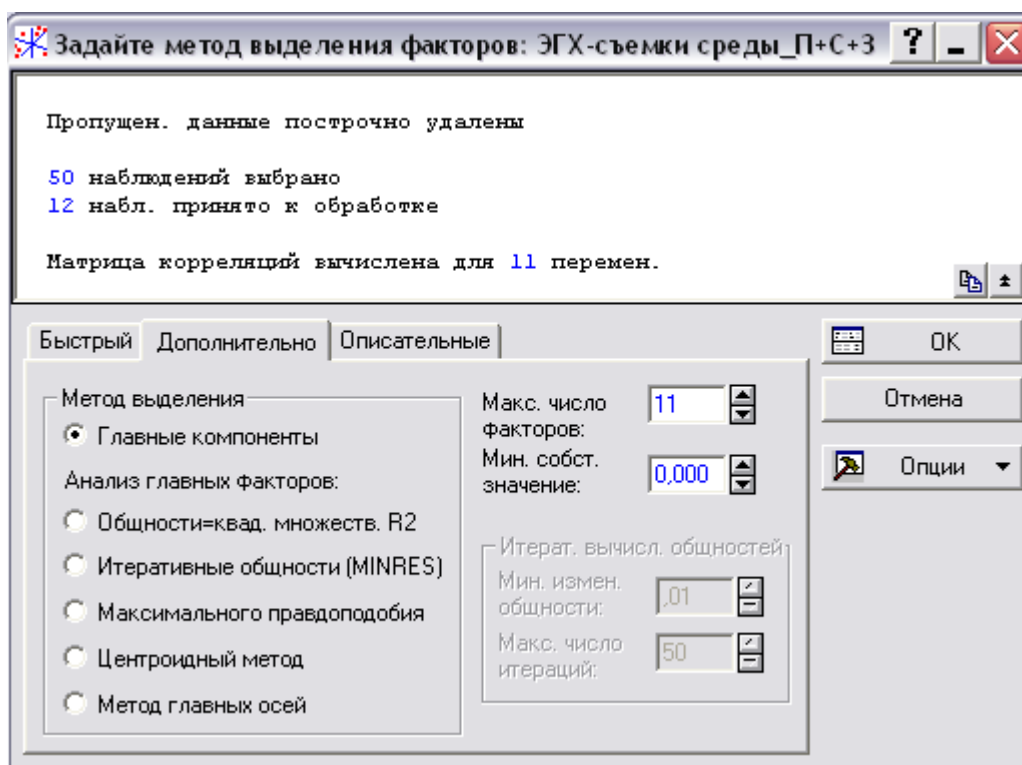



Рис. 3.109. Режим **Дополнительно** для выбора метода анализа

Собственные значения меньше указанного игнорируются. В окне **Метод выделения** выберите в качестве метода анализа **Главные компоненты** и щелкните **OK**. На экране появятся результаты анализа (рис. 3.110).

В верхней части окна **Результаты факторного анализа** дается информационное сообщение, в частности, приводятся значения собственных значений: 5,36454 2,46787 1,24187 ,808842 ,463248 ...

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

Шаг 6. Например, для просмотра расширенной информации о собственных значениях в режиме **Объясненная дисперсия** (рис.3.110) щелк-

ните по кнопке  Собственные значения. Дисперсия, объясненная последовательностью собственных значений, предстанет в форме рис. 3.111.

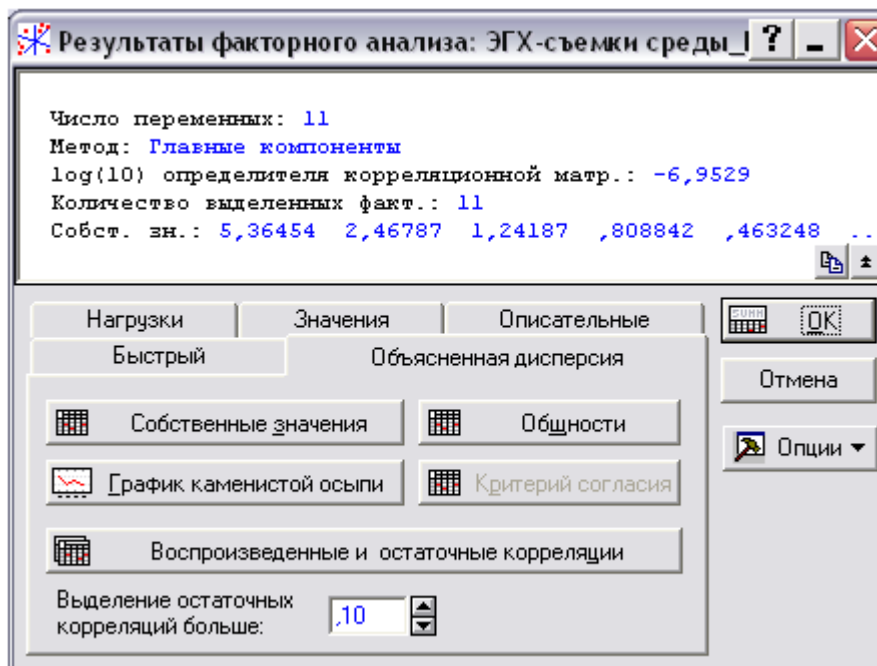


Рис. 3.110. Результаты факторного анализа для выборки С после применения метода главных компонент

Значен.	Собственные значения (ЭГХ-съемки среды_П+С+3)			
	Соб. зн.	% общей дисперс.	Кумулятивн. собст. знач.	Кумулятивн. %
1	5,364540	48,76854	5,36454	48,7685
2	2,467867	22,43515	7,83241	71,2037
3	1,241866	11,28969	9,07427	82,4934
4	0,808842	7,35311	9,88311	89,8465
5	0,463248	4,21135	10,34636	94,0578
6	0,329125	2,99205	10,67549	97,0499
7	0,157998	1,43635	10,83349	98,4862
8	0,084499	0,76818	10,91799	99,2544
9	0,059406	0,54006	10,97739	99,7945
10	0,018951	0,17228	10,99634	99,9668
11	0,003657	0,03325	11,00000	100,0000

Рис. 3.111. Дисперсия, объясненная последовательностью всех собственных значений

В первом столбце (рис. 3.111) приведены собственные значения, во втором - % от общей дисперсии для каждого собственного значения, в третьем – накопленные собственные значения и в четвертом – их % от общей дисперсии. Таким образом, первое собственное значение объясняет $\approx 48,8\%$ от общей дисперсии, совокупность первых двух собственных значений – уже $\approx 71,2\%$, а совокупность первых трех $\approx 82,5\%$ от общей дисперсии, что соответствует критерию объяснения главными компонентами не менее 80% всей дисперсии.

Согласно критерию Kaiser (Kaiser, 1960), следует сохранить собственные значения, больше чем 1. В соответствии с рис. 3.111, этот критерий предложил бы ограничиться первыми 3-мя собственными значениями.

Для проведения графического «теста каменной осыпи» Каттелла (Каттелл, 1966) выполните следующий шаг:

Шаг 7. В режиме Объясненная дисперсия (рис. 3.110) щелкните по кнопке График каменной осыпи для просмотра графической информации о всех собственных значениях (рис. 3.112).

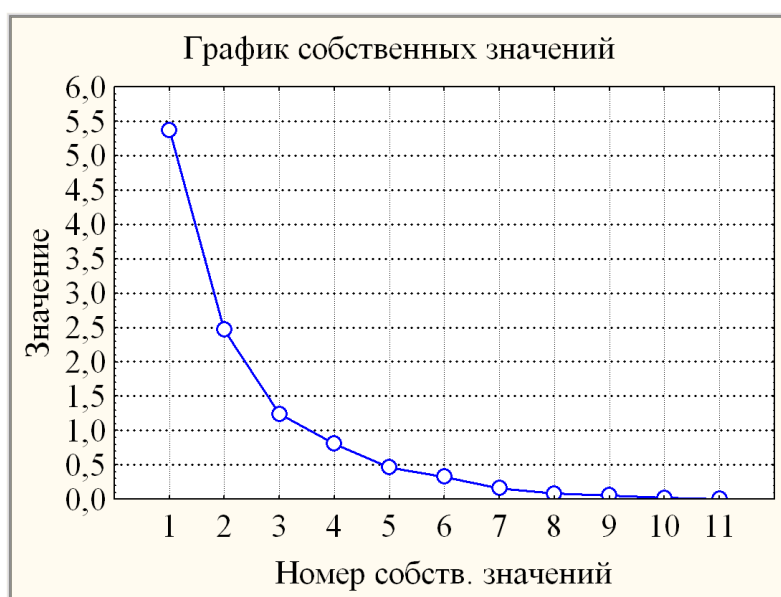
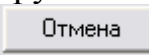

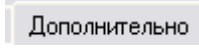

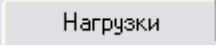


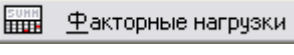


Рис. 3.112. График собственных значений

Рис. 3.112 демонстрирует графический метод («критерий каменной осыпи») оценки значимости собственных значений: справа от критической точки, где убывание собственных значений слева направо максимально замедляется, находится только «факториальная осыпь», которой пренебрегают. В соответствии с этим критерием можно оста-

вить 3 или 4 фактора. Поэтому, рекомендуется пробовать оба решения и выбрать вариант, поддающемуся лучшему толкованию.

Самую важную информацию, на которой базируется интерпретация факторов, содержат факторные нагрузки, которые могут интерпретироваться как корреляции между факторами и переменными. Посмотрим на вращаемые факторные нагрузки в 4-х факторной модели:

Шаг 8. Щелкните кнопкой , чтобы возвратиться в окно  (рис. 3.109) в режиме , в котором в поле справа укажите Максимальное число факторов, равное 4 (в соответствие с результатами кластеризации микроэлементов на рис. 3.98) и щелкните . На экране появятся результаты анализа 4-х факторной модели (рис. 3.113), В режиме  (рис. 3.113) выберите   и щелкните по кнопке  для просмотра факторных нагрузок в выбранной повернутой 4-х факторной модели (рис. 3.114).

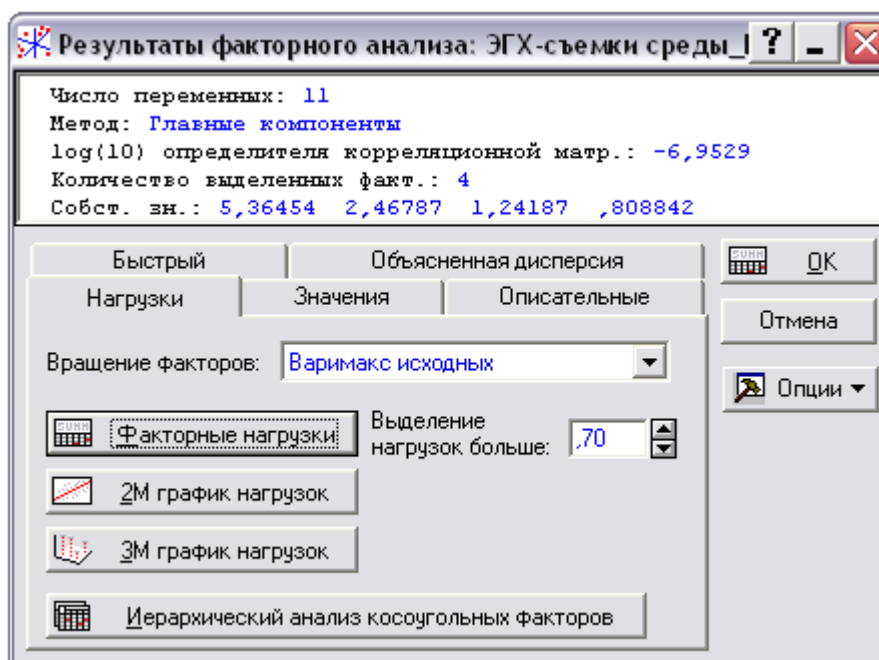


Рис. 3.113. Режим  в окне 

Если учесть, что основные компоненты извлечены так, чтобы последовательные компоненты объясняют всё меньшую дисперсию, то не

удивительно видеть, что первый фактор показывает большинство самых высоких (по модулю) нагрузок.

Перемен.	Фактор.нагрузки (Варим. исх.) (ЭГХ-съемки среды_П+С+З) Выделение: Главные компоненты (Отмечены нагрузки >,700000)			
	Фактор 1	Фактор 2	Фактор 3	Фактор 4
Pb	0,017	0,883	0,095	0,199
Zn	0,860	0,304	0,189	0,260
Cr	0,596	-0,090	-0,200	0,669
Ni	0,807	0,374	0,382	0,162
Co	0,476	0,180	0,794	0,227
Cu	0,585	-0,120	0,059	0,608
Mo	0,226	0,218	0,100	0,923
Mn	0,955	0,068	-0,083	0,229
V	0,271	0,876	0,042	-0,060
Sr	-0,112	0,116	0,949	-0,098
Ba	0,226	0,767	0,519	0,079
Общ. дис.	3,417	2,486	2,054	1,926
Доля общ	0,311	0,226	0,187	0,175

Рис. 3.114. Вращаемые факторные нагрузки в выбранной 4-х факторной модели

Для предания факторам простой интерпретации пространство факторов подвергнуто вращению так, что факторы отмечены высокими нагрузками для одних переменных, низкими нагрузками для других, и при этом сохраняется немного высоких поперечных нагрузок, то есть, немного переменных с существенными нагрузками на больше чем одном факторе. Самый стандартный вычислительный метод вращения, чтобы вызвать простую структуру – варимакс-вращение (Kaiser, 1958).

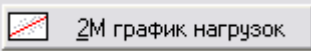
Согласно рис. 3.114, расположенные в порядке убывания высокие факторные нагрузки микроэлементов распределились по факторам следующим образом:

Фактор Ф1 – наиболее весомый, составной, характеризуется в основном микроассоциацией **Mn – Zn– Ni**, а также в меньшей степени **Cr– Cu** и **Co**, связанных с Ф1 положительной корреляционной связью.

Фактор Ф2 – менее весомый, составной, характеризуется в основном микроассоциацией **Pb–V–Ba**, связанной с Ф2 положительной корреляционной связью.

Фактор Ф3 – еще менее весомый, составной, характеризуется в основном микроассоциацией **Sr– Co**, а также в меньшей степени **Ba**, связанных с Ф3 положительной корреляционной связью.

Фактор Ф4 – еще менее весомый, составной, характеризуется в основном микроассоциацией **Mo– Cr– Cu**, связанной с Ф4 положительной корреляционной связью.

Шаг 9. Для геометрической интерпретации факторной модели щелкните, например, по кнопке  и выберите факторы для графика: Ф1 и Ф2 (рис. 3.115).

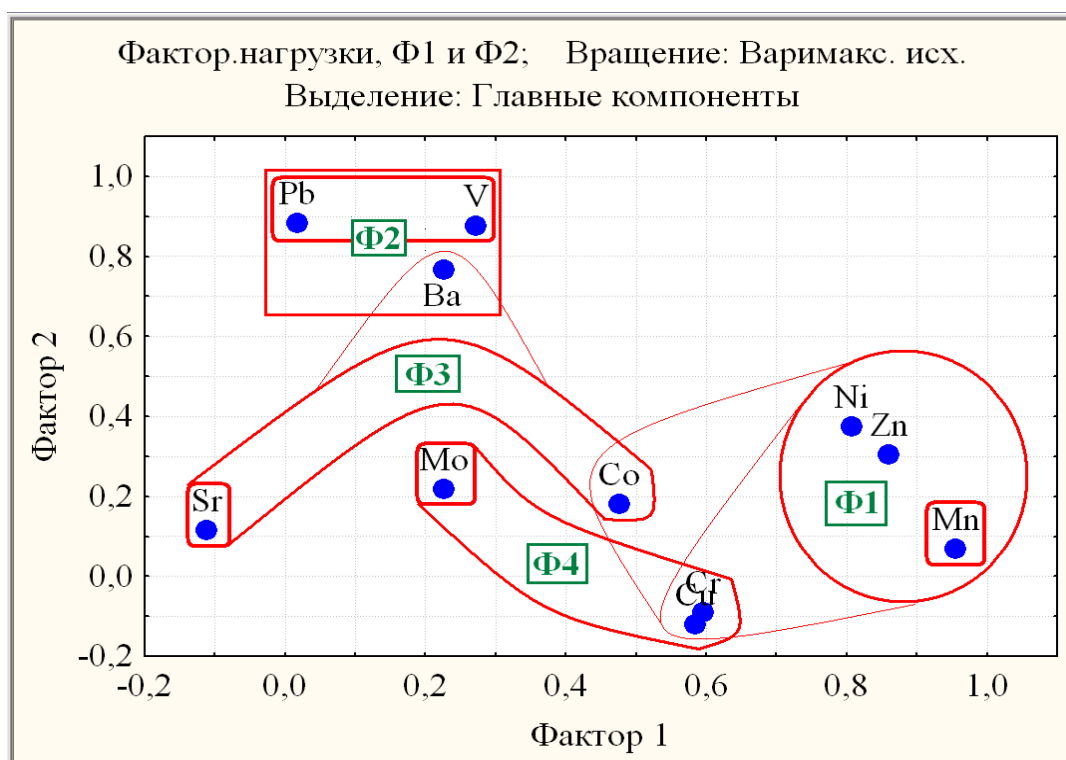
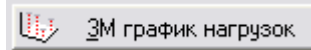


Рис. 3.115. Двумерный график факторных нагрузок микроэлементов выборки С

Диаграмма на рис. 3.115 наглядно иллюстрирует две микроассоциации **Mn – Zn– Ni** (Ф1) и **Pb–V–Ba** (Ф2) в координатах {Ф1, Ф2}, а также еще проекции двух групп **Sr– Co** (Ф3) и **Mo– Cr– Cu** (Ф4) на плоскость координат {Ф1, Ф2}. Последние две микроассоциации могут быть более наглядными в координатах {Ф1, Ф2, Ф3} или {Ф1, Ф2, Ф4},

доступных с помощью . На рис. 3.115' изображена 3-х мерная диаграмма рассеяния факторных нагрузок микроэлементов выборки С в координатах {Ф1, Ф2, Ф4}, достаточно наглядно разделяющая 11 микроэлементов на 4-е микроассоциации.

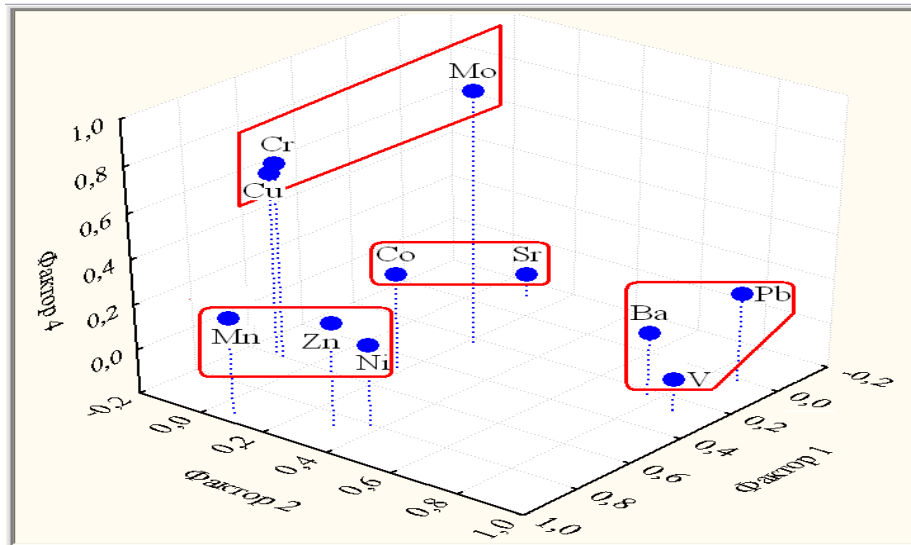


Рис. 3.115'. Диаграмма рассеяния факторных нагрузок микроэлементов выборки С в координатах {Ф1, Ф2, Ф4}

Для проведения дополнительного способа проверки достаточного числа выделенных факторов посмотрите воспроизведенную корреляционную матрицу, которая должна быть близка исходной.

Шаг 10. В полученном ранее окне **Результаты факторного анализа** В уже знакомом режиме **Объясненная дисперсия** (рис. 3.110) щелкните по кнопке **Воспроизведенные и остаточные корреляции** (рис. 3.116)

Перем	Воспроизведенные корреляции (ЭГХ-съемки среды_П+С+З)										
	Выделение: Главные компоненты										
	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
Pb	0,83	0,35	0,04	0,41	0,29	0,03	0,39	0,11	0,77	0,17	0,75
Zn	0,35	0,94	0,62	0,92	0,67	0,64	0,52	0,89	0,49	0,09	0,55
Cr	0,04	0,62	0,85	0,48	0,26	0,75	0,71	0,73	0,03	-0,33	0,01
Ni	0,41	0,92	0,48	0,96	0,79	0,55	0,45	0,80	0,55	0,30	0,68
Co	0,29	0,67	0,26	0,79	0,94	0,44	0,44	0,45	0,31	0,70	0,68
Cu	0,03	0,64	0,75	0,55	0,44	0,73	0,67	0,68	0,02	-0,08	0,12
Mo	0,39	0,52	0,71	0,45	0,44	0,67	0,96	0,43	0,20	0,00	0,34
Mn	0,11	0,89	0,73	0,80	0,45	0,68	0,43	0,98	0,30	-0,20	0,24
V	0,77	0,49	0,03	0,55	0,31	0,02	0,20	0,30	0,85	0,12	0,75
Sr	0,17	0,09	-0,33	0,30	0,70	-0,08	0,00	-0,20	0,12	0,94	0,55
Ba	0,75	0,55	0,01	0,68	0,68	0,12	0,34	0,24	0,75	0,55	0,91

Рис. 3.116. Воспроизведенная корреляционная матрица выборки С

Сравнение исходной наблюдаемой (рис. 3.108) и воспроизведенной (рис. 3.116) корреляционных матриц выборки S на примере пары $P_b - V$ можно оформить аналогично рис. 3.90:

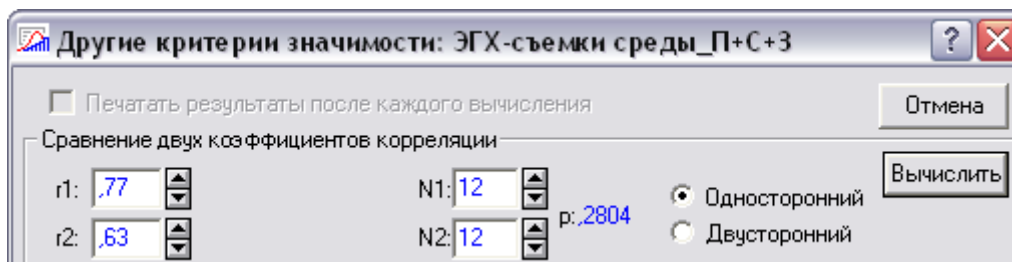


Рис. 3.117. Сравнение исходного и воспроизведенного коэффициентов корреляции пары $P_b - V$

Уровень значимости $p = 0,2804 > 0,10$ соответствует незначимому отличию. Таким образом, 4-х факторная модель обеспечивает воспроизведение исходных переменных с незначимыми отличиями. Заметим, что согласно рис. 3.111 первые четыре собственных значения объясняют $\approx 89,8\%$ от общей дисперсии.

Для проведения дальнейшего статистического анализа выборки S в рамках построенной 4-х факторной модели следует сохранить значения наблюдений в новой системе факторных координат.

Шаг 11. В окне **Результаты факторного анализа** войдите в режим **Значения** (рис. 3.118) и щелкните по кнопке **Сохранить значения** для сохранения значений факторов по каждому наблюдению (рис. 3.119)

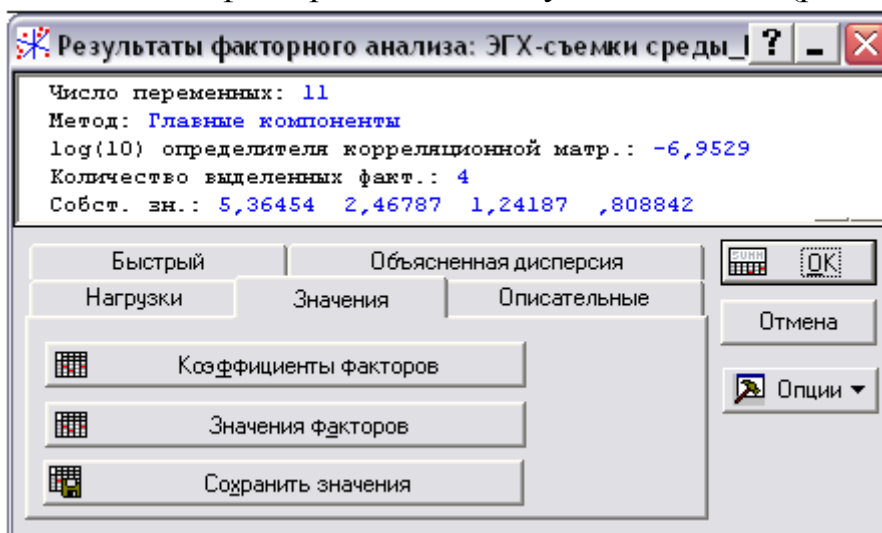


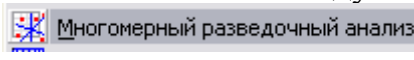
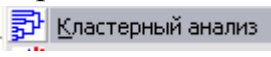


Рис. 3.118. Режим **Значения** в окне **Результаты факторного анализа:**


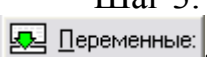

Набл.	Значения факторов (ЭГХ-съемки среды_П+С+З)			
	Фактор 1	Фактор 2	Фактор 3	Фактор 4
34	-1,36727	-0,14211	0,57446	-0,08724
35	-1,43890	-0,51072	0,89598	0,69726
36	-0,34546	-1,24091	0,97492	-1,10059
37	-0,11905	-0,72998	-0,87537	0,27077
38	1,95345	-0,51441	0,90946	-0,88289
39	0,13549	0,48867	1,08913	1,44362
40	1,42314	-0,91416	-0,72689	1,90053
41	0,61898	2,45145	0,81210	-0,34540
42	-0,19301	0,12411	0,12095	-0,27787
43	-0,02776	0,68373	-1,53252	-0,47404
44	-0,78952	0,77197	-1,31124	0,34180
45	0,14991	-0,46763	-0,93097	-1,48594

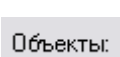
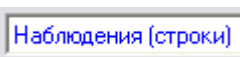
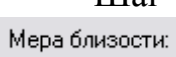

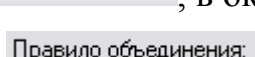


Рис. 3.119. Значения новых (факторных) переменных по каждому наблюдению для 4-х факторной модели выборки С

3.2.7. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов.

Шаг 1. Запустите в головном меню модуль .

Шаг 2. Выберите  в стартовой панели модуля и далее щелкните мышкой на разновидности . В появившемся окне выделите разновидность  и щелкните  (рис. 3.95).

Шаг 3. В открывшемся окне  нажмите кнопку , выберите все факторные переменные (рис. 3.120) и нажмите .

Шаг 4. Далее в окне  выделите , в окне  выделите , а в окне  —  (рис. 3.120) и нажмите .

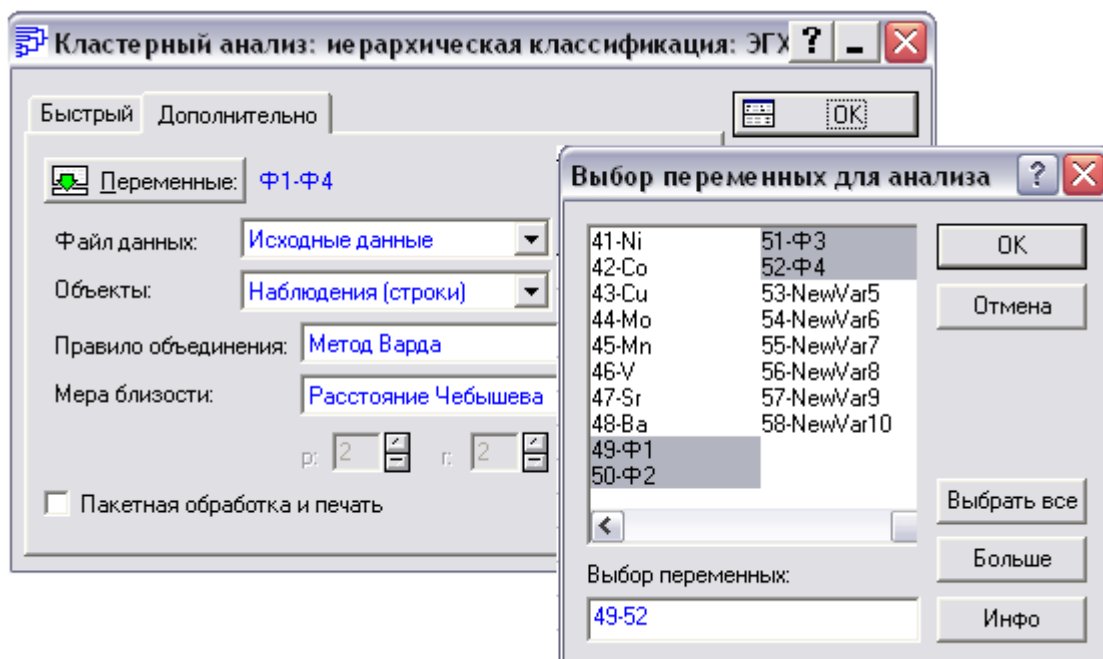


Рис.3.120. Окно **Кластерный анализ** для наблюдений:

Иерархическая классификация

Шаг 5. После этого щелкните **Вертикальная дендрограмма** (рис. 3.121).

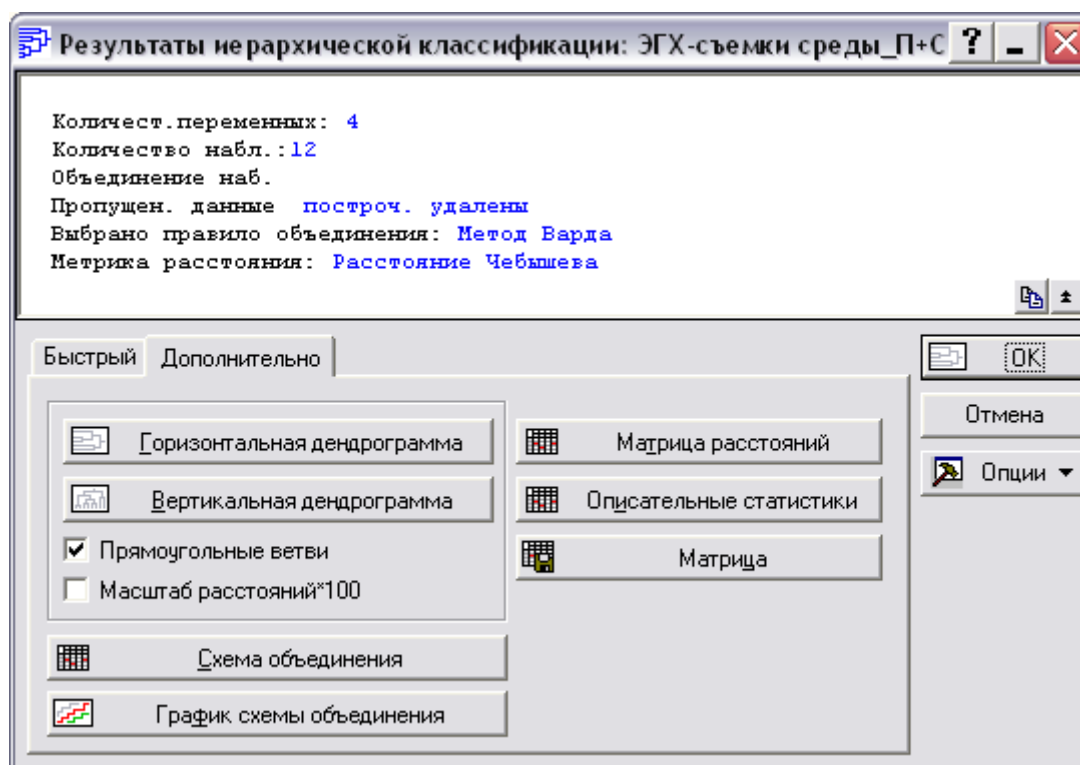


Рис. 3.121. Окно **Результаты иерархической классификации**:

Следующая дендрограмма наблюдений появится на экране:

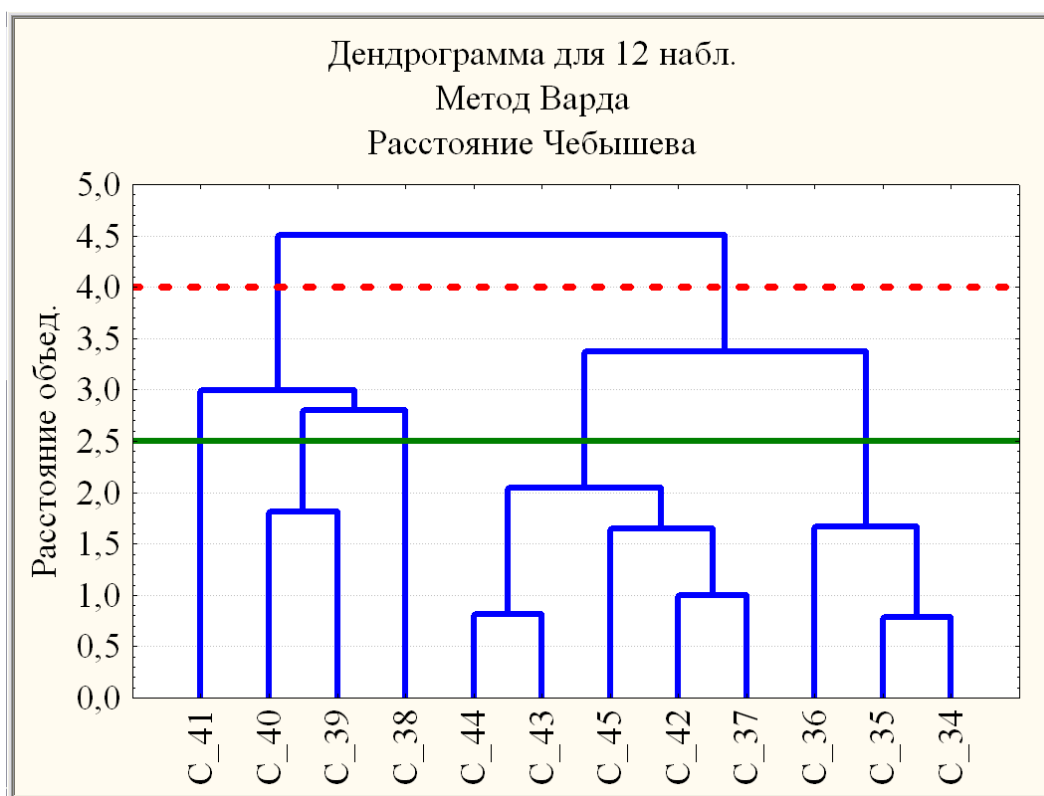









Рис. 3.122. Дендрограмма наблюдений в пространстве {Ф1, Ф2, Ф3, Ф4}

Таким образом, в построенном 4-х мерном факторном пространстве {Ф1, Ф2, Ф3, Ф4} проведена кластеризация наблюдений. При этом в качестве меры близости выбрано расстояние Чебышева, а в качестве правила объединения двух кластеров использован дисперсионный по сути метод Уорда (Варда). Методом древовидной кластеризации построено иерархическое дерево (рис. 3.122). В зависимости от выбора расстояния объединения можно получить соответствующее число кластеров. Так, например, расстоянию объединения, равному 4,0 (пунктирная горизонтальная прямая) соответствует 2 кластера, а расстоянию объединения, равному 2,5 (сплошная горизонтальная прямая) – 5 кластеров. Выбор значения связующего расстояния позволяет проводить кластеризацию на любом уровне, то есть строить кластерную модель с любым наперед заданным числом кластеров. При этом, увеличивая расстояния объединения, можно отслеживать динамику объединения кластеров по мере уменьшения степени их однородности.

Наряду с методом древовидной кластеризации применяется также метод *K*-средних, проводящий классификацию объектов (наблюдений) по заданному количеству кластеров. Алгоритм метода *K*-средних,

перемещая объекты в разные кластеры с целью минимизации изменчивости внутри кластеров и максимизации изменчивости между кластерами, оценивает качество кластеризации наблюдений по каждому фактору посредством параметрического дисперсионного анализа.

Шаг 6. В окне  **Методы кластеризации:** (рис. 3.95) выберите метод  **Кластеризация методом К-средних** и щелкните .

Шаг 7. В открывшемся окне  **Кластерный анализ:** нажмите кнопку  **Переменные:**, выберите все факторные переменные аналогично рис. 3.120 и нажмите . В окне **Объекты:** выделите **Наблюдения (строки)**, в окне **Число кластеров:** выделите 5 в соответствии с рис. 3.122 (сплошная горизонтальная прямая), а в окне **Начальные центры кластеров** — **Сортировать расстояния** (рис.3.123) и нажмите  для просмотра результатов кластеризации наблюдений методом К-средних (рис. 3.124)

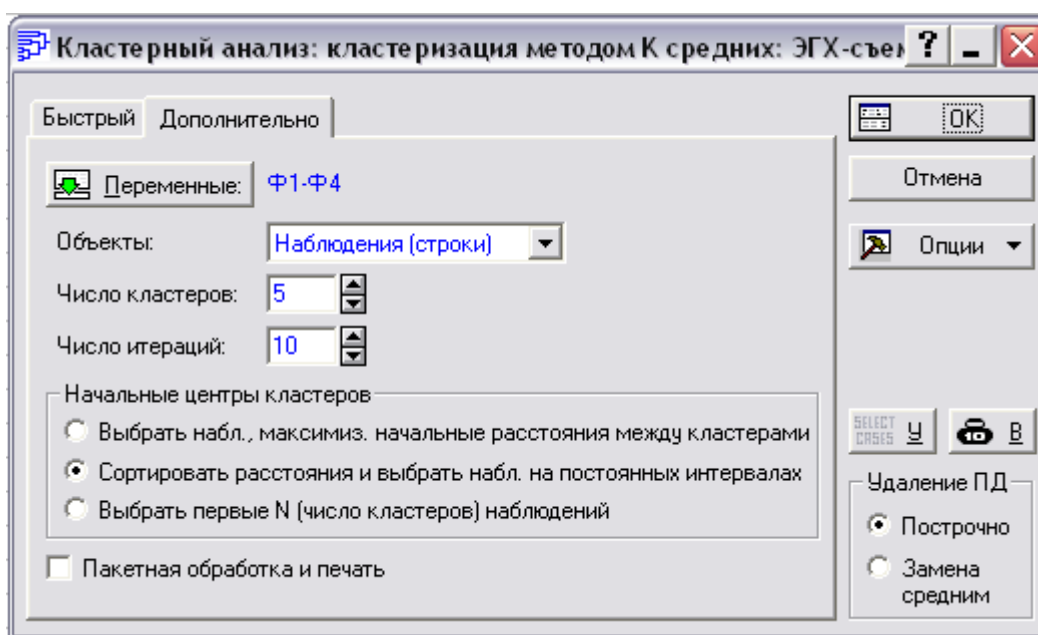





Рис. 3.123. Окно  **Кластерный анализ:** для наблюдений выборки С: 

Метод К-средних является итеративной процедурой. Алгоритм, реализованный в  **Кластерный анализ:**, весьма эффективен и обычно нет необходимости менять установку по умолчанию (10 итераций) в поле ввода **Число итераций:**.

Опции в окне **Начальные центры кластеров** позволяют выбрать способ вычисления начального расположения центров кластеризации, от которого до некоторой степени зависят результаты метода К-средних, если имеется много небольших кластеров, которые ясно различимы. Опция **«Максимизация расстояний между кластерами»** выбирает перебором наблюдения в качестве начального расположения кластеров так, чтобы максимизировать начальные кластерные расстояния. Заметим, что процедура может дать кластеры с единственным наблюдением, если они явно выделяются среди данных. Если выбрана опция **«Сортировать расстояния и взять наблюдения на постоянных интервалах»**, то сначала сортируются расстояния между всеми объектами, а затем в качестве начальных кластерных центров выбираются объекты на постоянных интервалах. Опция **«Выбрать первых N (количество кластеров) наблюдений»** выбирает первые N (количество кластеров) наблюдений в качестве начальных кластерных центров. Тем самым она даёт пользователю полный контроль над выбором начальной конфигурации, позволяя перемещать в начало файла наблюдения, которые вы хотите выбрать в качестве начальных кластерных центров.

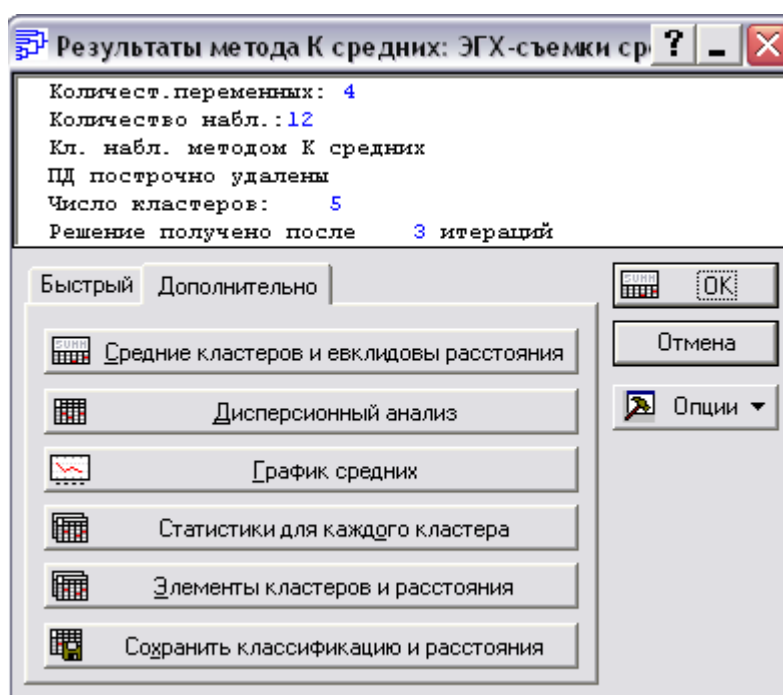
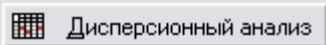


Рис. 3.124. Окно **Результаты метода К средних** для выборки С.

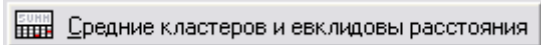
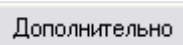
Рассмотрим результаты кластеризации наблюдений методом К-средних (рис. 3.124)

Шаг 8. Нажмите прежде всего кнопку  для построения стандартной таблицы результатов дисперсионного анализа кластеризации наблюдений по каждому фактору (рис. 3.125).

перемен.	Дисперсионный анализ (ЭГХ-съемки среды_П+С+3)					
	Между SS	сс	Внутри SS	сс	F	значим. p
Ф1	8,917	4	2,083	7	7,489	0,01135
Ф2	7,589	4	3,411	7	3,894	0,05668
Ф3	7,648	4	3,352	7	3,993	0,05363
Ф4	7,099	4	3,901	7	3,185	0,08656

Рис. 3.125. Результаты дисперсионного анализа кластеризации наблюдений по факторам

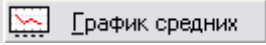
Признаком хорошей классификации является малость изменчивости внутри кластера в сравнении с изменчивостью вне кластера, что осуществляется путём стандартного межгруппового дисперсионного анализа для каждого фактора. Значения F -статистики, полученные для каждого фактора, является индикатором того, насколько хорошо соответствующий фактор дискриминирует кластеры. Значения p -уровней определяют статистически значимым (на уровне значимости $0,005 < p \approx 0,011 < 0,050$) различия между кластерами для фактор Ф1 и слабо значимым (на уровне значимости $0,05 < p < 0,10$) различия между кластерами для остальных факторов (Ф2, Ф3 и Ф4), используемых в анализе.

Шаг 9. Нажмите кнопку  в режиме  для вывода на экран таблицы результатов со средними для каждого кластера по каждому фактору (рис. 3.126).

перемен.	Средн.класт. (ЭГХ-съемки среды_П+С+3)				
	Кластер № 1	Кластер № 2	Кластер № 3	Кластер № 4	Кластер № 5
Ф1	0,619	-1,051	-0,196	0,779	1,953
Ф2	2,451	-0,631	0,076	-0,213	-0,514
Ф3	0,812	0,815	-0,906	0,181	0,909
Ф4	-0,345	-0,164	-0,325	1,672	-0,883

Рис. 3.126. Кластерные средние в факторном пространстве

Здесь же доступна просмотру таблица результатов с евклидовыми расстояниями между "центрами кластеров". В данном случае матрица показывает евклидовы расстояния между кластерами, вычисленные из средних значений соответствующих кластеров по факторам, используемым для классификации.

Шаг 9'. Нажмите кнопку  для вывода на экран графиков факторных средних для каждого кластера (рис. 3.127).

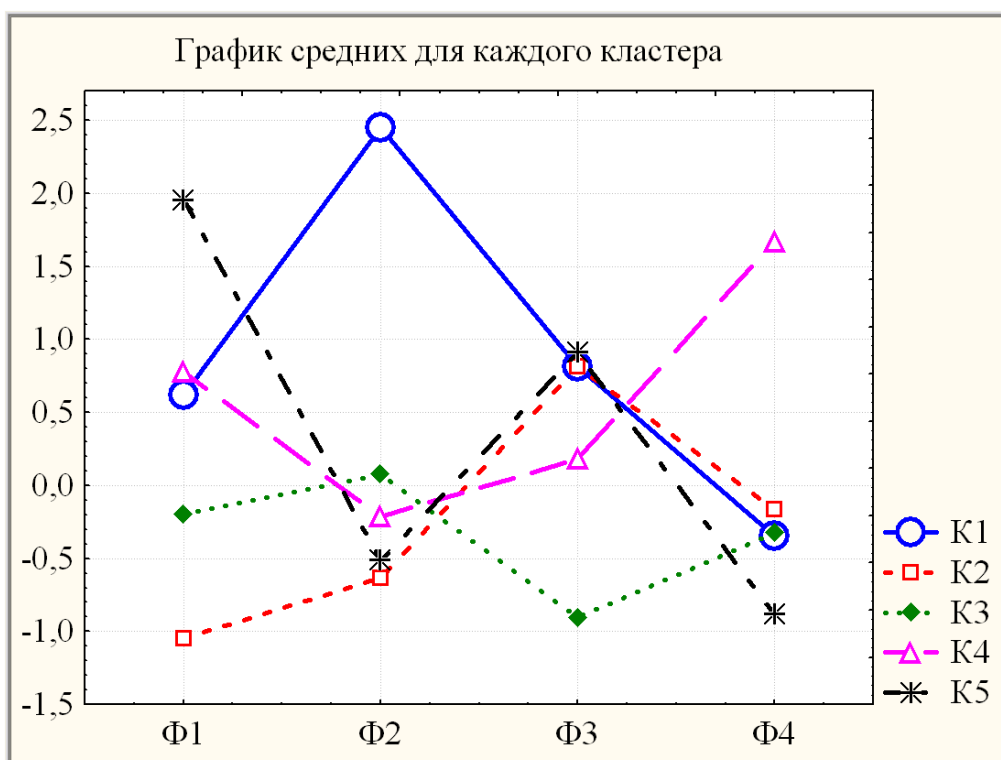
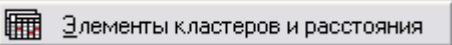



Рис. 3.127. Линейные графики факторных средних для каждого кластера

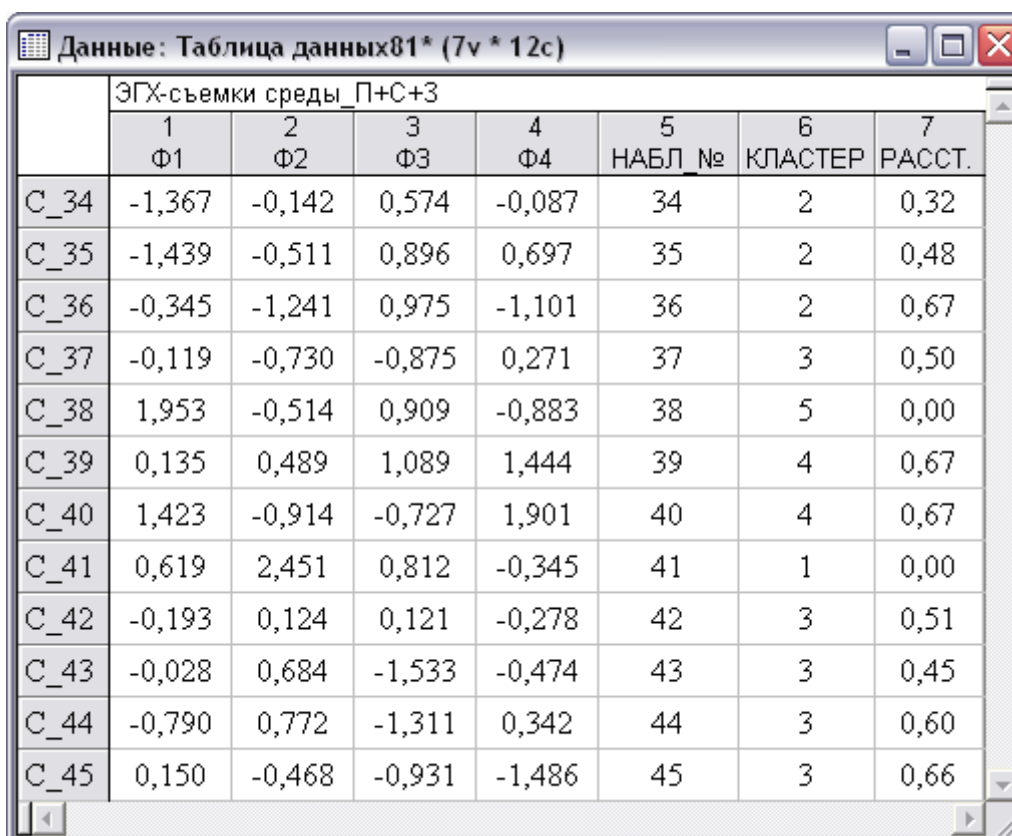
Шаг 10. Нажмите  для вывода на экран списка элементов (наблюдений) каждого кластера, где для каждого наблюдения вычислено его расстояние до центра кластера (рис. 3.128).

Элементы кластера номер 3 (ЭГХ-съемки среды_П+С+З) и расстояния до центра кластера. Кластер содержит 5 набл.					
	Набл. С_37	Набл. С_42	Набл. С_43	Набл. С_44	Набл. С_45
Расст.	0,503	0,514	0,451	0,601	0,664

Рис. 3.128. Элементы кластера K3 с расстояниями до центра

Полученные результаты кластеризации наблюдений методом *K*-средних (с опцией «Сортировать расстояния и взять наблюдения на постоянных интервалах»,) в полной мере соответствуют иерархическому дереву (рис. 3.122), построенному методом древовидной кластеризации, в котором в качестве меры близости выбрано евклидово расстояние, а в качестве правила объединения двух кластеров использован метод Уорда (Варда).

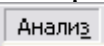
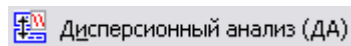
Шаг 10'. Нажмите  для вывода на экран данных, содержащий информацию о классификации и расстояниях наблюдений до центра соответствующего кластера, в дополнение к факторным переменным, выбранным из исходного файла данных (рис. 3.129).

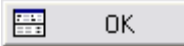
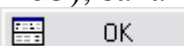


	ЭГХ-съемки среды П+С+З						
	1 Ф1	2 Ф2	3 Ф3	4 Ф4	5 НАБЛ_№	6 КЛАСТЕР	7 РАССТ.
C_34	-1,367	-0,142	0,574	-0,087	34	2	0,32
C_35	-1,439	-0,511	0,896	0,697	35	2	0,48
C_36	-0,345	-1,241	0,975	-1,101	36	2	0,67
C_37	-0,119	-0,730	-0,875	0,271	37	3	0,50
C_38	1,953	-0,514	0,909	-0,883	38	5	0,00
C_39	0,135	0,489	1,089	1,444	39	4	0,67
C_40	1,423	-0,914	-0,727	1,901	40	4	0,67
C_41	0,619	2,451	0,812	-0,345	41	1	0,00
C_42	-0,193	0,124	0,121	-0,278	42	3	0,51
C_43	-0,028	0,684	-1,533	-0,474	43	3	0,45
C_44	-0,790	0,772	-1,311	0,342	44	3	0,60
C_45	0,150	-0,468	-0,931	-1,486	45	3	0,66

Рис. 3.129. Результаты факторной классификации наблюдений

Предварительно оформим стандартную процедуру дисперсионного анализа результатов факторной классификации наблюдений.

Шаг 11. Запустите  и щелкните  (см. рис. 3.64). В появившемся окне выберите тип анализа (однофакторный дисперсионный анализ) и задайте метод (диалог быстрых специфика-

ций), как показано на рис.3.64, и нажмите . В появившемся окне выберите в качестве зависимых все факторные переменные (позиции 49-52), а в качестве категориальной КЛАСТЕР (позиция 53), закажите все коды (1-5), как показано на рис. 3.130, и нажмите .

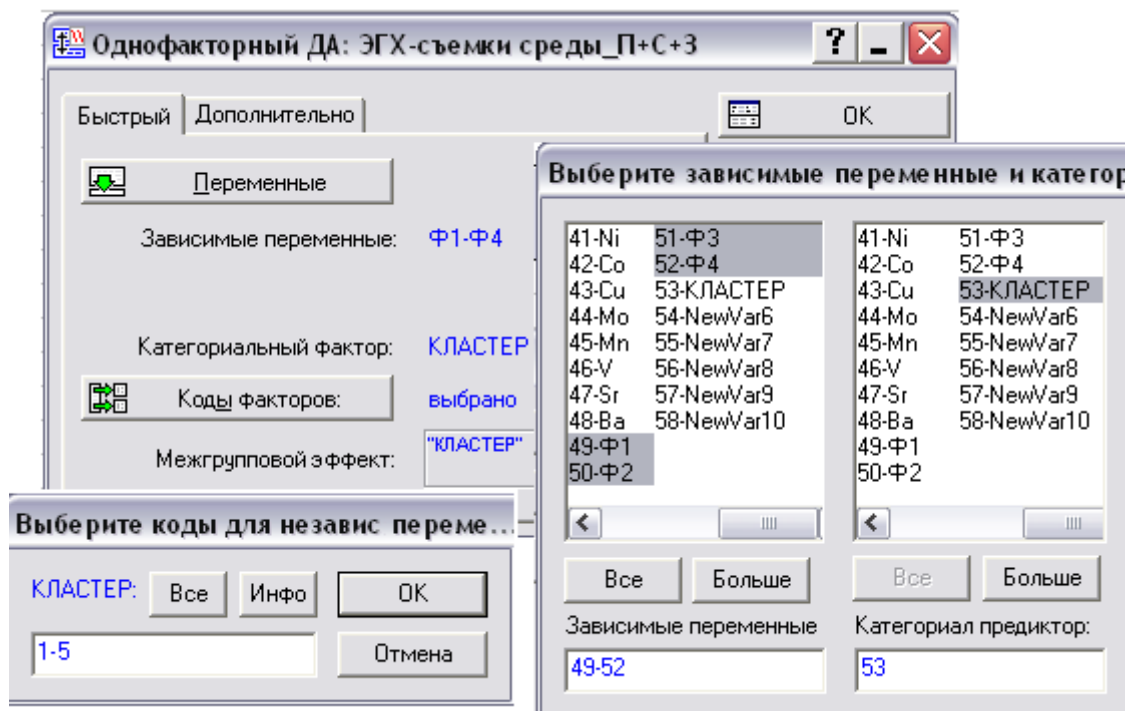

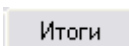
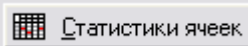
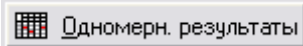
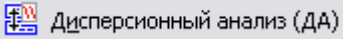


Рис. 3.130. Выбор переменных для дисперсионного анализа кластеров

Шаг 12. После этого в появившемся окне  (рис.3.66) в режиме  нажмите  для просмотра результатов описательной статистики (расширенная таблица средних на рис. 3.126) и  для просмотра результатов дисперсионного анализа (рис. 3.125).

Продемонстрируем возможности модуля  для расширенного анализа полученного результата. Так как значения одномерного F -критерия для каждого фактора – значимы (рис. 3.125), то далее можно использовать метод множественных сравнений для выявления тех групп кластеров, которые вызвали данный эффект и максимально отличаются от других, а также для каждого фактора (рис. 3.129) аналогично рис. 3.71-72 можно выделить однородные (различающиеся незначимо, то есть $p > 0,10$) иногда пересекающиеся группы кластеров.

Шаг 13. В окне **Результаты анализа** (рис. 3.66) нажмите кнопку **Больше** и в открывшемся окне выберите режим **Апостериорные**, в котором представлены различные апостериорные критерии (рис. 3.131). Выберите, например, критерии **Фишера НЗР**, **Тьюки** и **Шеффе** с опцией отображения **Однородные группы:** (рис. 3.132).

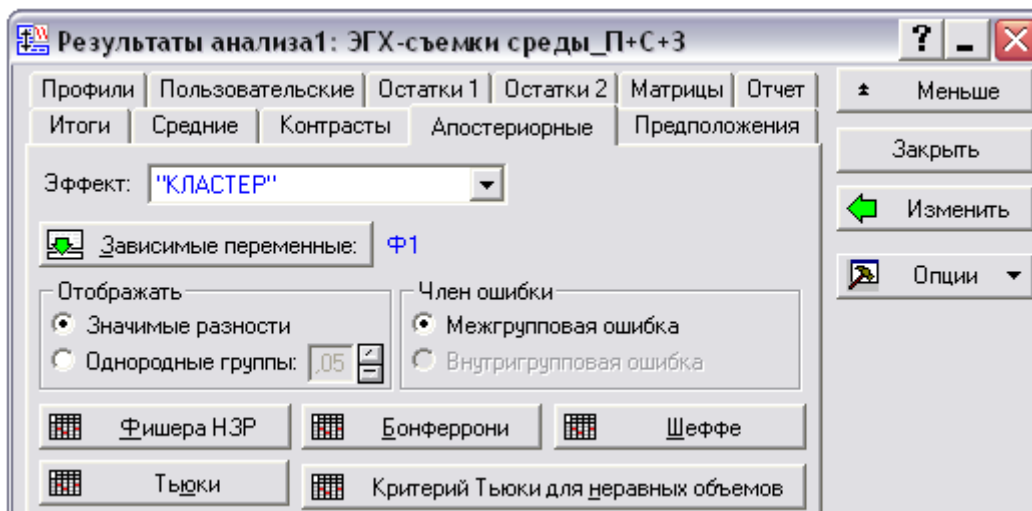


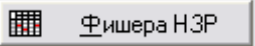
Рис. 3.131. Режим **Апостериорные** для выбора апостериорных критериев множественных сравнений средних



Phi1	1	2	3
Среднее			
2	-1,051		***
3	-0,196	***	***
1	0,619	***	***
4	0,779	***	***
5	1,953		***

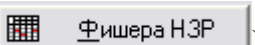
Phi1	1	2	3
Среднее			
2	-1,051	***	
3	-0,196	***	***
1	0,619	***	***
4	0,779		***
5	1,953		***

Phi1	1	2
Среднее		
2	-1,051	***
3	-0,196	***
1	0,619	***
4	0,779	***
5	1,953	***

Рис. 3.132. Результаты сравнений кластерных средних для $\Phi 1$ по критериям **Фишера НЗР**, **Тьюки** и **Шеффе**

Согласно рис. 3.132, для фактора Φ_1 по апостериорному критерию  выделены 3 последовательно пересекающиеся однородные группы кластеров, расположенные в порядке убывания средних: $\{K_5, K_4, K_1\}, \{K_4, K_1, K_3\}, \{K_3, K_2\}$. Разделение на 3 группы кластеров является статистически значимым ($0,05 > p > 0,005$). Так, например, K_5 ($m_{K_5} \approx 1,953$) отличается от K_3 ($m_{K_3} \approx -0,196$) на уровне значимости $p \approx 0,009$, а K_1 ($m_{K_1} \approx 0,619$) отличается от K_2 ($m_{K_2} \approx -1,051$) на уровне значимости $p \approx 0,033$.

Остальные апостериорные критерии подтверждают этот результат в несколько смягченной форме: по критерию  выделены также 3 однородные группы кластеров: $\{K_5, K_4, K_1\}, \{K_4, K_1, K_3\}, \{K_1, K_3, K_2\}$, а по критерию  – только 2-е пересекающиеся группы: $\{K_5, K_4, K_1, K_3\}, \{K_4, K_1, K_3, K_2\}$. Таким образом, критерий НЗР дает самые значимые результаты по сравнению с другими апостериорными критериями.

Аналогичным образом (на основании рис. 3.132 для Φ_1 по критерию ) можно оформить результаты сравнений кластерных средних для остальных факторов. Таким образом:

Φ_1 : $\{K_5, K_4, K_1\}, \{K_4, K_1, K_3\}, \{K_3, K_2\}$ так, что K_5 ($m_{K_5} \approx 1,953$) отличается от K_3 ($m_{K_3} \approx -0,196$) на уровне значимости $p \approx 0,009$, а K_1 ($m_{K_1} \approx 0,619$) отличается от K_2 ($m_{K_2} \approx -1,051$) на уровне значимости $p \approx 0,033$.

Φ_2 : $\{K_1\}, \{K_3, K_4, K_5, K_2\}$, так, что K_1 ($m_{K_1} \approx 2,451$) отличается от K_3 ($m_{K_3} \approx 0,076$) на уровне значимости $p \approx 0,017$.

Φ_3 : $\{K_5, K_2, K_1, K_4\}, \{K_1, K_4, K_3\}$ так, что K_2 ($m_{K_2} \approx 0,815$) отличается статистически значимо от K_3 ($m_{K_3} \approx -0,906$) на уровне значимости $p \approx 0,011$.

Φ_4 : $\{K_4\}, \{K_2, K_3, K_1, K_5\}$ так, что K_4 ($m_{K_4} \approx 1,672$) отличается от K_2 ($m_{K_2} \approx -0,164$) на уровне значимости $p \approx 0,031$.

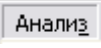
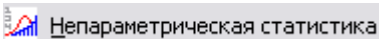
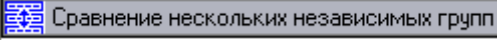
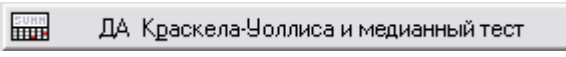
Результаты кластерного анализа наблюдений по совокупности факторов (рис. 3.126-127) с учетом результатов сравнений кластерных средних для каждого фактора позволяют провести классификацию наблюдений в порядковой шкале стандартизированных измерений, полагая в качестве уровня «Средний» – стандартизированный интервал $(-0,5; +0,5)$ наблюдений, «Выше среднего» – $(+0,5; +1,5)$ наблюдений, «Ниже среднего» – $(-1,5; -0,5)$. Аномально высокие значения $(> +1,5)$ определяют уровень «Лидер», а аномально низкие значения $(< -1,5)$ определяют уровень «Аутсайдер».

Таблица 3.21

Классификация наблюдений по совокупности факторов в порядковой шкале стандартизированных измерений.

Кластер	Уровень кластера на фоне среднего по фактору				Состав кластера
	Ф1 (Mn–Zn–Ni)	Ф2 (Pb–V–Ba)	Ф3 (Sr – Co)	Ф4 (Mo–Cr–Cu)	
К1	Выше среднего	Лидер	Выше среднего	Средний	C_41
К2	Ниже среднего	Ниже среднего	Выше среднего	Средний	C_34+35+36
К3	Средний	Средний	Ниже среднего	Средний	C_37+42+ +43+44+45
К4	Выше среднего	Средний	Средний	Лидер	C_39+40
К5	Лидер	Средний – Ниже среднего	Выше среднего	Ниже среднего	C_38

Для достоверности полученных результатов дисперсионного анализа необходимо проверить предположения о нормальном распределении сравниваемых кластеров и об однородности дисперсий в них. Но в данном случае достаточно замечания о малочисленности состава кластеров, чтобы обратиться к непараметрическому однофакторному дисперсионному анализу на основе критерия Краскела-Уоллиса и медианного теста.

Шаг 14. Запустите модуль  и щелкните в нем по кнопке  (см. рис. 3.133). Затем выберите сравнение нескольких независимых групп  (рис. 3.133), закажите переменные стандартным образом (см. рис. 3.130) и щелкните по кнопке  (рис. 3.134). В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа на основе критерия Краскела-Уоллиса и медианного теста (рис. 3.135).

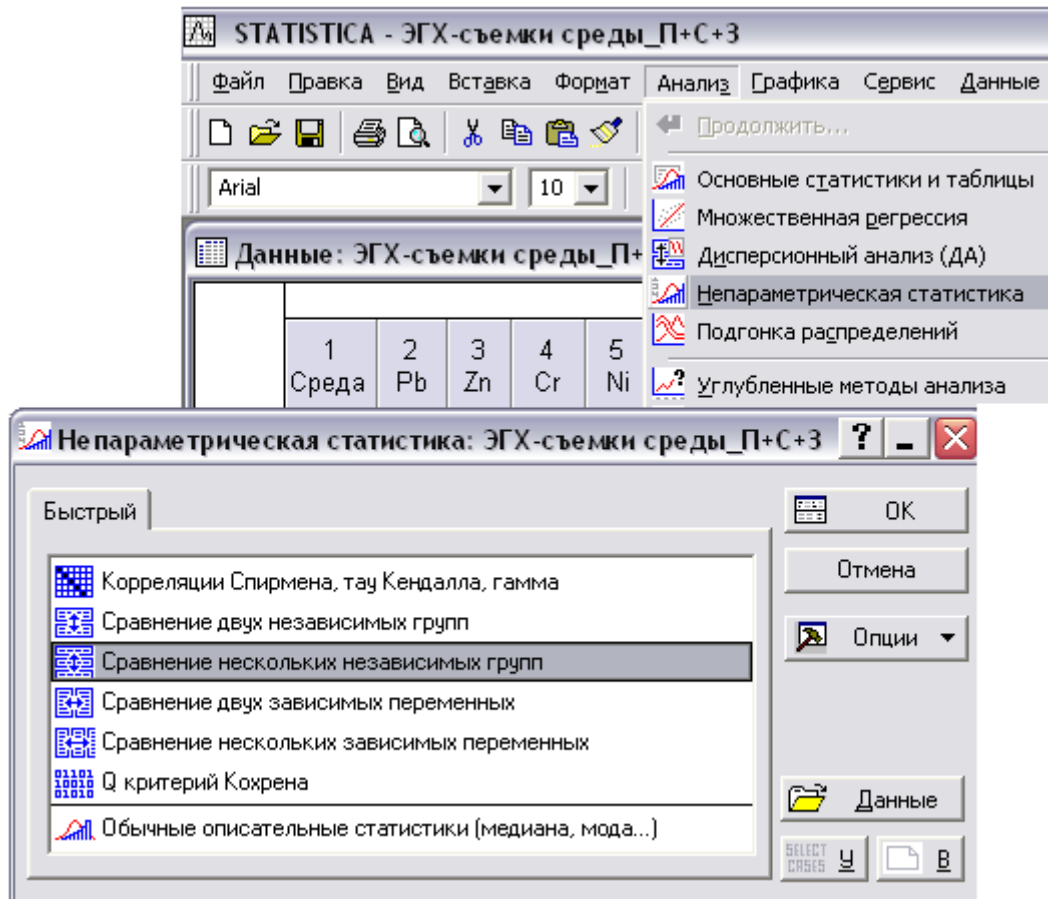


Рис. 3.133. **Непараметрическая статистика** в головном модуле **Анализ** для базы данных **ЭГХ-съемки среды_П+С+З**

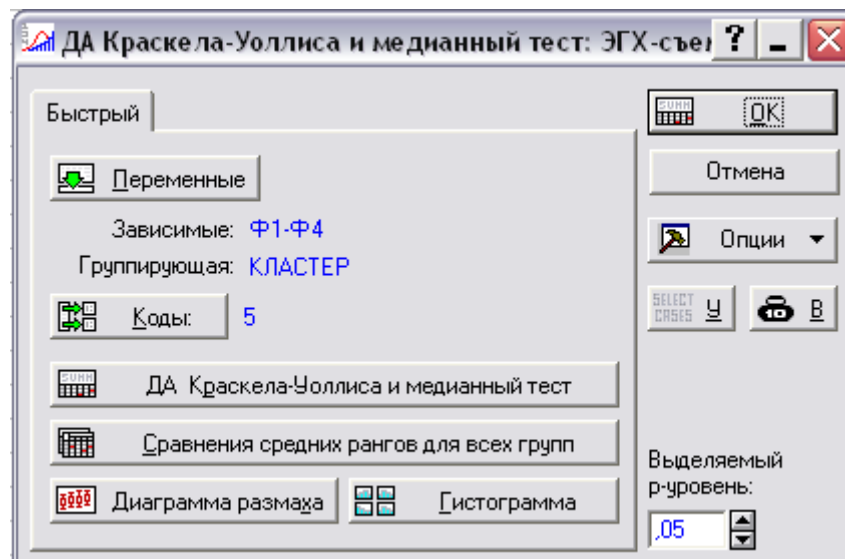


Рис.3.134. Окно непараметрического однофакторного ДА

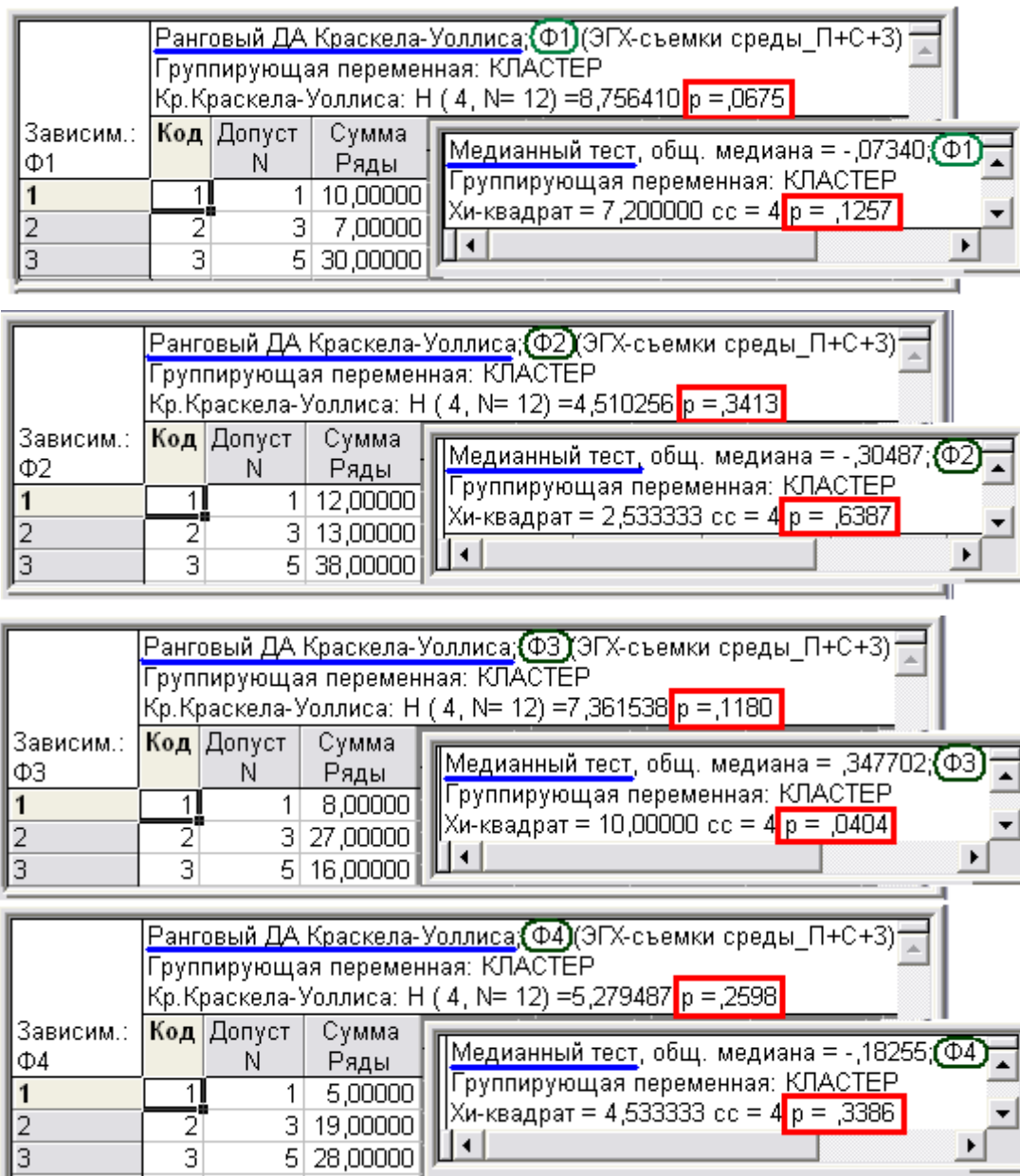


Рис. 3.135. Результаты рангового дисперсионного анализа кластеризации наблюдений С на основе критерия Краскела-Уоллиса и медианного теста

По совокупности критерия Краскела-Уоллиса и медианного теста (рис. 3.135) 5 кластеров К1-К5 различается слабо значимо в случае Ф1 и Ф3 ($0,10 > p > 0,05$), но незначимо в случае Ф2 и Ф4 ($p > 0,10$). Таким образом, непараметрический дисперсионный анализ подтверждает (в смягченной форме) результаты параметрического дисперсионного анализа о значимой кластеризации наблюдений выборки С только в случае факторов Ф1 и Ф3. В силу расхождения результатов непараметрического и параметрического однофакторного дисперсионного анализа в связи

с малочисленностью состава кластеров более корректными признаются результаты непараметрического дисперсионного анализа.

ВЫВОДЫ

➤ Согласно табл. 3.14, высоко значимыми по совокупности трех сред являются различия содержаний следующих химических элементов Pb ($m_C = 99,17$; $m_{II} = 22,58$ и $m_3 = 3,56$), Cr ($m_{II} = 159,39$; $m_C = 69,17$ и $m_3 = 16,00$) и Co ($m_{II} = 9,76$; $m_C = 5,17$ и $m_3 = 1,60$); незначимыми по совокупности трех сред являются различия средних содержаний Mn ($m_{II} = 412,12$; $m_C = 325,0$ и $m_3 = 448,0$). Остальные химические элементы имеют смешанные варианты различия содержаний по средам: например, содержания Zn в почве ($m_{II} = 73,64$) и в снеговом осадке ($m_C = 82,50$) различаются незначимо, в то же время средние содержания Zn в почве ($m_{II} = 73,64$) и в золе растений ($m_3 = 140,8$) различаются высоко значимо, а средние содержания Zn в снеговом осадке ($m_C = 82,50$) и в золе растений ($m_3 = 140,8$) различаются статистически значимо. В данном случае следует отметить специфику форм нахождения элементов в данных средах. Контрастность элементов в твердом осадке снега объясняется пробоподготовкой. В данном случае из большой массы пробы снега получают небольшую долю твердого вещества, которая значительно обогащена элементами.

➤ Как следует из табл. 3.16 – 18, наиболее значимые корреляционные связи в ассоциациях наблюдаются между Ni и Co в почве, между Ni и Zn в снеговом осадке, между V и Cr в золе растений. Присутствие значимой связи в почве между Ni и Co объясняется присутствием в пробах ферромагнетита – продукта литейного производства. Микросферы магнитные черного цвета и обогащены элементами группы железа. Данные образования широко распространены на территориях с присутствием чугуно-литейного производства.

➤ Из рис. 3.97-99, геохимические спектры микроэлементов в почве, снеговом осадке и золе растений существенно различаются. В данном случае может быть уловлена закономерность только в том случае, когда пробы отбираются в одних точках, а это в условиях городской территории сделать затруднительно.

➤ Построена 4-х факторная модель выборки C : Ф1 (Mn – Zn – Ni), Ф2 (Pb – V – Ba), Ф3 (Sr – Co) и Ф4 (Mo – Cr – Cu), объясняющая $\approx 90\%$ от общей дисперсии.

➤ На основе 4-х мерного факторного пространства {Ф1, Ф2, Ф3, Ф4} построена 5-ти кластерная модель наблюдений.

3.3. Статистический анализ (СА) содержаний микроэлементов в накипи разных территорий, полученных методами ИНАА и ISP

Проведем сравнительный анализ содержаний 6-ти химических элементов в накипи населенных пунктов Томской (Т) и Челябинской (Ч) областей, полученных методами ИНАА (И) и ISP (I), на основе категориальной базы эколого-геохимических измерений, приведенной в приложении П₄. При этом, в зависимости от количества категорий (факторов) и их уровней, возможны разные формы сравнительного статистического анализа содержаний каждого из 6-ти микроэлементов исходной базы данных: парный сравнительный анализ, двухфакторный или однофакторный дисперсионные анализы. Рассмотрим их последовательно.

3.3.1. Парный сравнительный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.

В данном подходе каждому из 6-ти микроэлементов соответствует 4 отдельных по территориям и методам выборки (рис. 3.136): Т*И, Т*I, Ч*И и Ч*I. Здесь, например, выборка Т*И определяет содержание химического элемента в накипи населенных пунктов Томской области, полученных методом ИНАА согласно приложению П₄.

	1 Na _{T*И}	2 Na _{T*I}	3 Na _{Ч*И}	4 Na _{Ч*I}	5 Ca _{T*И}	6 Ca _{T*I}	7 Ca _{Ч*И}	8 Ca _{Ч*I}
1	0,003	1,950	0,170	1,830	22,9	35,2	29,0	25,5
2	0,010	2,500	0,130	0,890	29,1	37,3	48,6	36,9
3	0,012	2,080	0,090	2,480	28,0	37,0	29,8	21,7
4	0,003	1,470	0,060	2,330	23,9	35,3	5,6	3,6
5	0,015	2,320	0,070	1,050	25,8	36,8	35,8	34,5
6	0,003	0,600	0,070	2,690	29,4	37,6	54,4	37,5
7	0,003	2,070	0,060	2,400	25,8	38,0	42,5	34,1
8	0,017	2,750	0,100	0,630	23,7	38,8	49,9	37,3
9			0,080	1,630			53,8	36,8
10			0,090	0,620			50,0	40,5
11			0,060	1,510			52,7	38,9
12			0,090	0,790			54,8	39,0
13			0,080	0,370			49,9	37,0
14			0,070	1,600			52,9	41,4
15			0,060	2,260			52,4	41,1

Рис.3.136. Фрагмент отдельной базы данных П₄.
Анализ в данном случае проводится аналогично п. 3.1.1-3.

Выборочные характеристики объем выборки (n), среднее (m), медиана (Me), стандартное отклонение (S), асимметрия (A), эксцесс (E) и их стандартные ошибки (σ_A , σ_E), а также минимум (**Min**) и максимум (**Max**) содержаний химических элементов (24 выборки) вычисляются аналогично рис. 3.1-2. Фрагмент приведен в табл. 3.22.

Таблица 3.22

Оценки числовых характеристик содержаний микроэлементов Na и Ca в выборках Т*И, Т*І, Ч*И и Ч*І.

	n	m	Me	Min	Max	S	A	σ_A	E	σ_E
Na _{Т*И}	8	0,008	0,007	0,003	0,017	0,006	0,41	0,75	-1,88	1,48
Na _{Т*І}	8	1,968	2,075	0,600	2,750	0,672	-1,26	0,75	1,87	1,48
Na _{Ч*И}	15	0,085	0,080	0,060	0,170	0,030	1,86	0,58	3,79	1,12
Na _{Ч*І}	15	1,539	1,600	0,370	2,690	0,777	-0,01	0,58	-1,49	1,12
Ca _{Т*И}	8	26,075	25,80	22,90	29,40	2,522	0,19	0,75	-1,72	1,48
Ca _{Т*І}	8	37,000	37,15	35,19	38,78	1,237	-0,38	0,75	-0,52	1,48
Ca _{Ч*И}	15	44,140	49,90	5,60	54,80	13,74	-1,86	0,58	3,50	1,12
Ca _{Ч*І}	15	33,715	37,03	3,57	41,42	9,985	-2,32	0,58	5,76	1,12

Последние 4 столбца используются для формирования критерия соответствия эмпирического распределения нормальному теоретическому в виде отношения коэффициентов асимметрии A и эксцесса E к их стандартным отклонениям σ_A и σ_E (\tilde{t}_1 и \tilde{t}_2 в (1.22)). При аналогичной проверке случайной величины X на логнормальный закон распределения по определению (1.11) формируется (например, в Excel) случайная величина $\ln X$, которая проверяется на нормальный закон. В связи с малыми объемами выборок рекомендуется принимать критическую величину для \tilde{t}_1 и \tilde{t}_2 равной 2 (Каждан и др., 1990).

Проверка гипотезы о нормальном (логнормальном) распределении содержаний микроэлементов (24 выборки) оформляется аналогично рис. 3.7-8.

В случае теста Колмогорова-Смирнова вместо неравенства можно вычислить соответствующий уровни значимости α_{k-s} по значению критерия d по формуле (1.21) аналогично рис.3.10.

Результаты проверки гипотезы о законе (нормальном или логнормальном) распределения содержаний микроэлементов (24 выборки) сведены в табл.3.23.

Таблица 3.23

Критерии (1.19, 21, 22) сравнения с N и LN распределений в П₄.

	α_{χ^2}		$\alpha_{K-S} (d)$		$\tilde{t}_1; \tilde{t}_2$	
	N	LN	N	LN	N	LN
Na_{T*И}	>0,10	>0,10	>0,10 (0,31)	>0,10 (0,32)	0,5; 1,3	0,2; 1,7
Na_{T*И}	>0,10	0,09	>0,10 (0,24)	>0,10 (0,31)	1,7; 1,3	1,2; 0,6
Na_{ч*И}	<10⁻³	0,04	>0,10 (0,24)	>0,10 (0,17)	3,2; 3,4	2,0; 1,1
Na_{ч*И}	>0,10	>0,10	>0,10 (0,16)	>0,10 (0,19)	0,01; 1,3	1,2; 0,6
Ca_{T*И}	>0,10	>0,10	>0,10 (0,18)	>0,10 (0,18)	0,3; 1,2	0,1; 1,2
Ca_{T*И}	>0,10	>0,10	>0,10 (0,20)	>0,10 (0,20)	0,5; 0,3	5,8; 11
Ca_{ч*И}	0,01	<10⁻³	>0,10 (0,29)	>0,10 (0,30)	3,2; 3,1	5,3; 9,3
Ca_{ч*И}	<10⁻³	<10⁻³	0,07 (0,32)	0,02 (0,38)	4,0; 5,1	5,8; 11
Fe_{T*И}	>0,10	>0,10	>0,10 (0,20)	>0,10 (0,17)	1,0; 0,3	0,8; 0,1
Fe_{T*И}	>0,10	>0,10	>0,10 (0,18)	>0,10 (0,16)	0,1; 1,0	0,2; 0,8
Fe_{ч*И}	>0,08	>0,10	>0,10 (0,22)	>0,10 (0,17)	1,5; 0,03	0,05; 0,4
Fe_{ч*И}	0,102	>0,10	>0,10 (0,28)	>0,10 (0,13)	2,3; 0,9	0,2; 0,8
Sr_{T*И}	>0,10	>0,10	>0,10 (0,23)	>0,10 (0,27)	1,8; 1,6	1,3; 1,7
Sr_{T*И}	>0,10	>0,10	>0,10 (0,25)	>0,10 (0,23)	2,2; 2,0	4,2; 6,4
Sr_{ч*И}	>0,10	<10⁻³	>0,10 (0,15)	>0,10 (0,25)	1,1; 1,2	3,7; 5,2
Sr_{ч*И}	>0,10	<10⁻³	>0,10 (0,17)	>0,10 (0,30)	0,9; 1,4	4,2; 6,4
Ba_{T*И}	0,06	>0,10	>0,10 (0,18)	>0,10 (0,12)	1,9; 1,8	0,1; 0,5
Ba_{T*И}	>0,10	>0,10	>0,10 (0,22)	>0,10 (0,22)	1,3; 1,3	4,2; 6,4
Ba_{ч*И}	0,05	>0,04	>0,10 (0,25)	>0,10 (0,19)	1,6; 0,03	2,0; 2,8
Ba_{ч*И}	>0,10	<10⁻³	>0,10 (0,20)	>0,10 (0,26)	0,8; 0,8	4,2; 6,2
Cr_{T*И}	0,08	>0,10	>0,10 (0,26)	>0,10 (0,16)	1,3; 0,4	0,3; 0,6
Cr_{T*И}	>0,10	>0,10	>0,10 (0,21)	>0,10 (0,25)	1,8; 1,6	2,4; 2,1
Cr_{ч*И}	>0,10	>0,10	>0,10 (0,16)	>0,10 (0,14)	0,6; 0,1	0,07; 0,16
Cr_{ч*И}	>0,10	>0,10	>0,10 (0,11)	>0,10 (0,18)	0,3; 0,7	2,4; 2,1

Применение критерия Пирсона (Каждан и др., 1990) предполагает разделение выборочных данных на k равномерных интервалов. Число k определяется с учетом эмпирической формулы $k \approx 1 + 4 \lg(n)$, что дает для объема выборки $n_T = 8$ значение $k \approx 4,6$ и для объема выборки $n_C = 15$ значение $k \approx 5,7$. С учетом чувствительности критерия Пирсона к разбиению выборки на k интервалов можно $\alpha(k)$ вычислять, например, по интерполяционной формуле Лагранжа (1.20) для двух целочисленных точек в случае выборки (Т) и двух или трех – в случае (Ч), ближайших к значению k , что в данном случае дает:

$$\alpha(4, 6) = 0,4\alpha_4 + 0,6\alpha_5 \text{ для выборки (Т) и}$$

$$\alpha(5, 7) = 0,3\alpha_5 + 0,7\alpha_6 \text{ или}$$

$$\alpha(5, 7) = 0,195\alpha_5 + 0,91\alpha_6 - 0,105\alpha_7 \text{ для (Ч).}$$

Сами $\alpha_4, \alpha_5, \alpha_6, \alpha_7$ (рис. 3.137 для нормального распределения $Na_{ч*И}$) вычисляются аналогично рис. 3.7-9, где с учетом малого объема выборок следует убрать галочку в окне теста хи - квадрат.

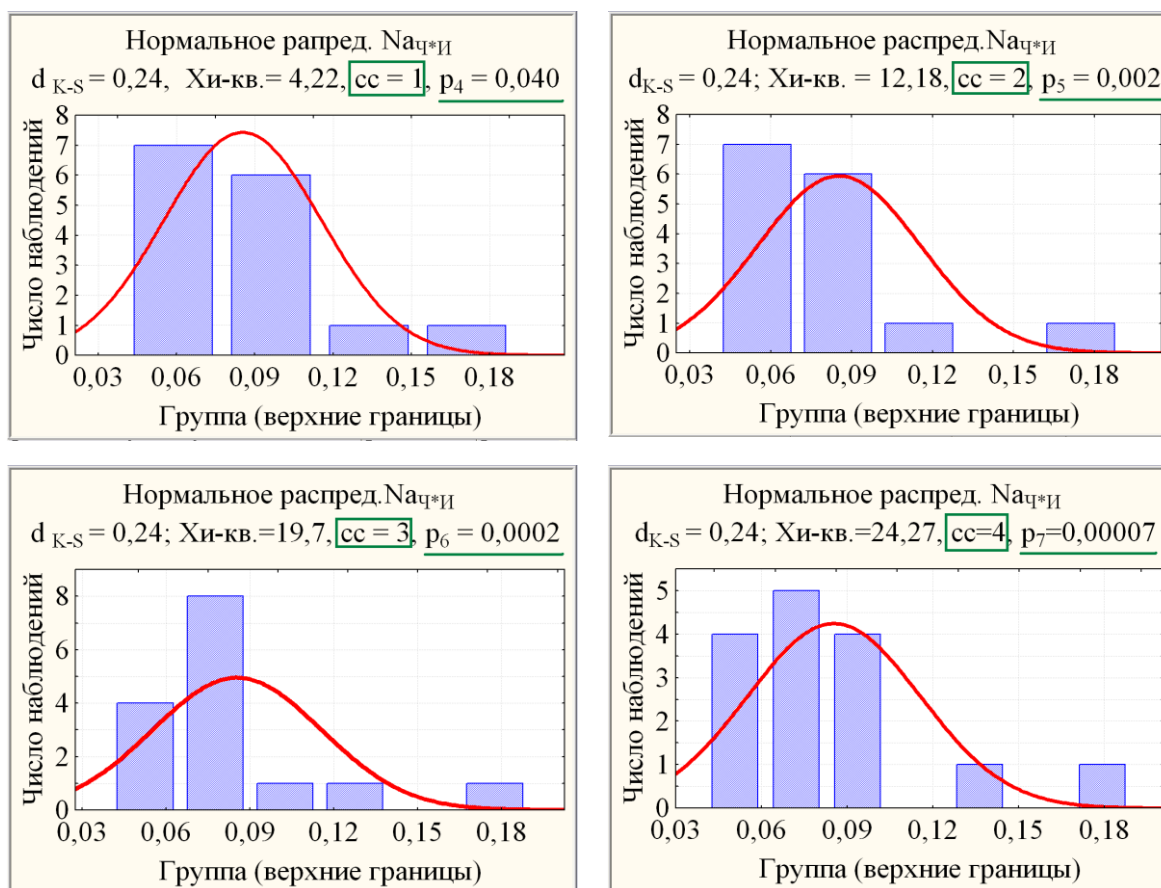


Рис. 3.137. Гистограмма распределения содержания $Na_{ч*И}$ с кривой плотности нормального распределения и результатами тестов Колмогорова-Смирнова (1.21) и Хи-квадрат (1.19).

Аналогично случаю нормального распределения (рис. 3.137) проверяется гипотеза о логнормальном распределении (рис. 3.138).

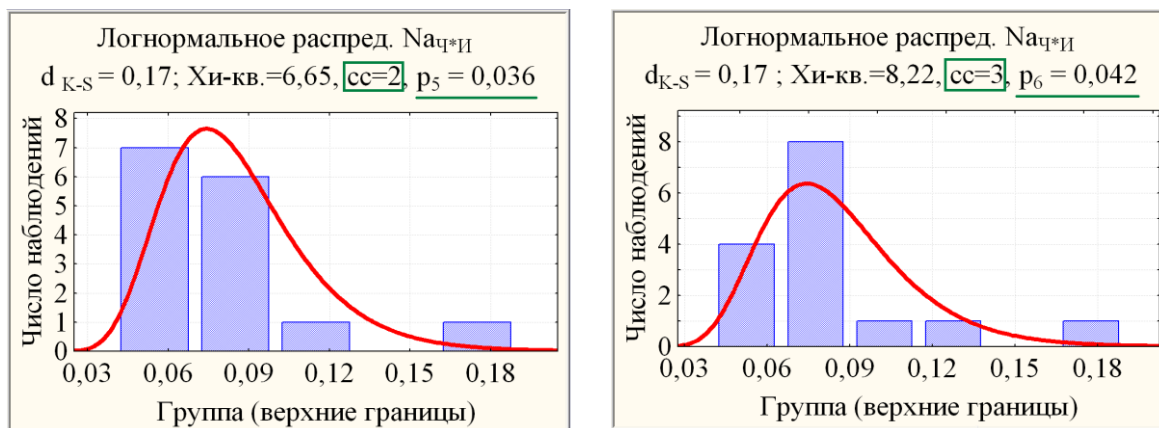


Рис. 3.138. Гистограмма распределения содержания $Na_{ч*И}$ с кривой плотности логнормального распределения и результатами тестов Колмогорова-Смирнова (1.21) и Хи-квадрат (1.19).

Все вспомогательные вычисления проводились в Excel (аналогично рис.3.10). Например, в случае сравнения с нормальным законом распределения содержания микроэлемента $Na_{ч*И}$ имеем (см. рис.3.33):

E1		fx = =0,195*B1+0,91*C1-0,105*D1			
	B	C	D	E	F
1	0,002	0,0002	0,00007	0,0006	

Рис. 3.139. Сглаживание по трем точкам, проводимое в Excel, зависимости уровня значимости $\alpha(k)$ критерия Пирсона от разбиения выборки на k интервалов

На рис.3.139 в ячейках B1, C1, D1 помещены соответственно значения $\alpha_5, \alpha_6, \alpha_7$, а в ячейке E1 значение $\alpha(5,7)$, вычисленное по интерполяционной формуле Лагранжа, отраженной в строке формул.

Согласно табл.3.23, по совокупности данных трех критериев нормально распределенными одновременно в двух городах можно считать содержания химических элементов Na_{ISp}, Fe, Sr, Ba, Cr , а логнормально распределенным – содержания химических элементов $Na_{ИНАА}$.

Результаты сравнения (по областям) средних содержаний химических элементов с учетом законов их распределения приведены в табл.3.24.

Проверка гипотезы о равенстве средних двух содержаний микроэлементов на основе выбранного закона распределения проводилась с

использованием параметрических критериев: t -критерия Стьюдента (уровень значимости α_t) для двух нормальных выборок с учетом F- критерия Фишера (уровень значимости α_F) аналогично рис.3.11-12 или критерия (1.26) Родионова (уровень значимости α_R) в случае логнормальной модели с учетом F-критерия Фишера (уровень значимости α_F) с помощью вычислений в Excel значений критерия и нахождения соответствующего значения уровня значимости с помощью калькулятора распределения вероятности аналогично рис.1.8 и 1.14. В противном случае применяется непараметрический двух выборочный критерий Колмогорова-Смирнова (уровень значимости α_{2K-S}) по аналогии с рис.3.13-14.

Таблица 3.24

Критерии (Стьюдента, Родионова и Колмогорова-Смирнова) сравнения средних содержаний микроэлементов с учетом законов их распределения в накипи из посуды населенных пунктов Т и Ч

	\bar{x}_T	$\bar{x}_Ч$	S_T	$S_Ч$	α_t (α_F)	α_R (α_F)	α_{2K-S}
Na_И	0,008	0,085	0,006	0,030		0,000 (0,00)	
Na_Г	1,968	1,539	0,672	0,777	0,20 (0,73)		
Ca_И	26,075	44,140	2,522	13,741			< .001
Ca_Г	37,000	33,715	1,237	9,985			> .10
Fe_И	0,679	0,667	0,493	0,303	0,95 (0,12)		
Fe_Г	0,716	0,201	0,392	0,202	0,000 (0,03)		
Sr_И	0,064	0,310	0,044	0,104	0,000 (0,03)		
Sr_Г	0,057	0,209	0,028	0,078	0,000 (0,02)		
Ba_И	0,025	0,064	0,014	0,040	0,003 (0,01)		
Ba_Г	0,017	0,008	0,009	0,004	0,003 (0,02)		
Cr_И	0,563	23,100	0,484	2,883	0,000 (0,00)		
Cr_Г	30,875	23,733	9,553	12,510	0,18 (0,48)		

Как следует из табл.3.24, к неоднородным (различия в средних – значимые) по областям следует отнести выборки Na_{II} , Ca_{II} , Fe_I , Sr , Ba и Cr_{II} . О значимом различии в содержаниях химического элемента Na_{II} по областям наглядно говорит визуальное сравнение соответствующих распределений на рис.3.140.)). Построение составных гистограмма распределения содержания микроэлементов проводится по аналогии с рис.3.15-16

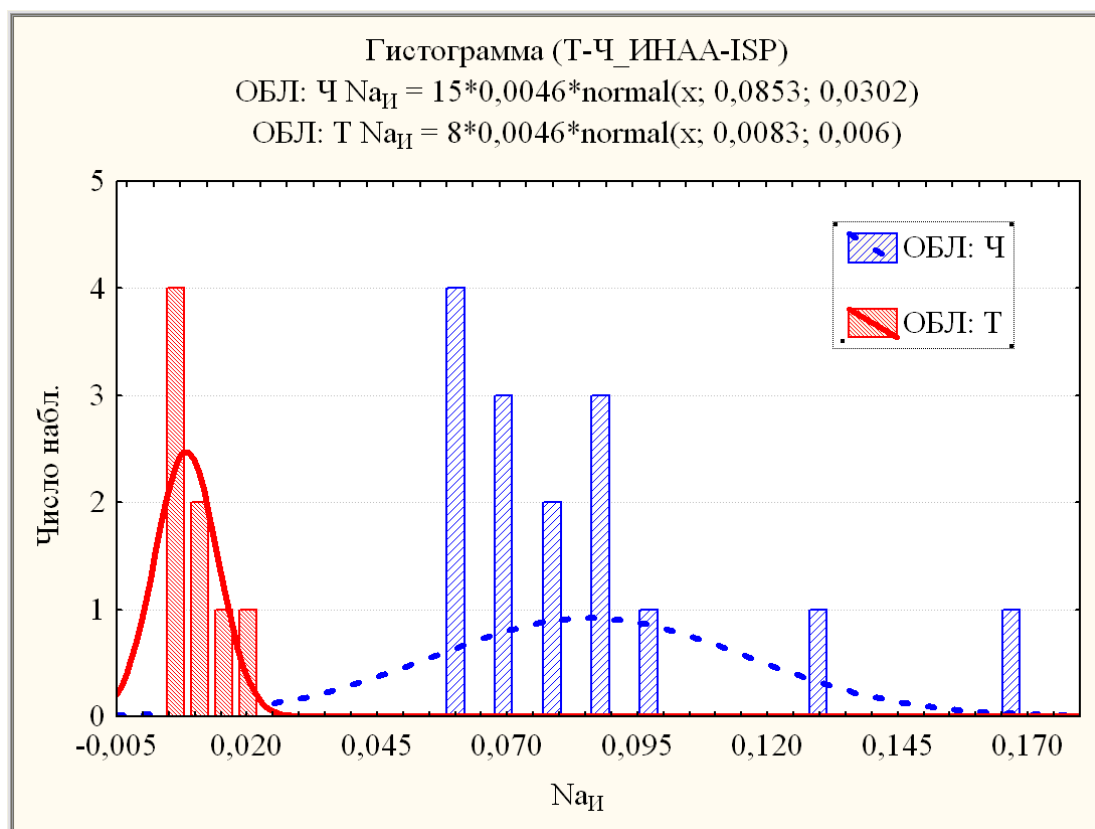


Рис. 3.140. Составная гистограмма логнормальных распределений содержания Na_{II} по областям Т и Ч

В неоднородных по областям выборках Na , Ca , Fe , Sr , Ba и Cr сравнения средних содержаний химических элементов в накипи по методам ИНАА и ISP проводились отдельно для Т (табл. 3.25) и Ч (табл.3.26). Причем, с учетом того, что согласно табл. 3.23 нормально распределенными являются только содержания химических элементов Fe , Sr , Ba и Cr , соответствующий t -критерий Стьюдента (α_t) с учетом F -критерия Фишера (α_F) применялся для Fe , Sr , Ba и Cr аналогично рис.3.11-12, а для остальных (Na и Ca) применялся аналогично рис.3.13-14 непараметрический двух выборочный критерий Колмогорова-Смирнова ($\alpha_{2К-С}$).

Таблица 3.25

Критерии сравнения средних содержаний микроэлементов в накипи населенных пунктов Т по методам ИНАА и ISP

Т						
	$\bar{x}_{\text{ИНАА}}$	\bar{x}_{ISP}	$S_{\text{ИНАА}}$	S_{ISP}	$\alpha_t(\alpha_F)$	α_{2K-S}
Na	0,008	1,968	0,006	0,672		< .001
Ca	26,075	37,000	2,522	1,237		< .001
Fe	0,679	0,716	0,493	0,392	0,869(0,560)	
Sr	0,064	0,057	0,044	0,028	0,723(0,290)	
Ba	0,025	0,017	0,014	0,009	0,196(0,209)	
Cr	0,563	30,875	0,484	9,553	0,000(0,000)	

Таблица 3.26

Критерии сравнения средних содержаний микроэлементов в накипи населенных пунктов Ч по методам ИНАА и ISP

Ч						
	$\bar{x}_{\text{ИНАА}}$	\bar{x}_{ISP}	$S_{\text{ИНАА}}$	S_{ISP}	$\alpha_t(\alpha_F)$	α_{2K-S}
Na	0,085	1,539	0,030	0,777		< .001
Ca	44,14	33,715	13,741	9,985		< .001
Fe	0,667	0,201	0,303	0,202	0,000 (0,144)	
Sr	0,310	0,209	0,104	0,078	0,006 (0,292)	
Ba	0,064	0,008	0,040	0,004	0,000 (0,000)	
Cr	23,100	23,733	2,883	12,510	0,850 (0,000)	

Согласно табл.3.25 (по Т) высоко значимо различаются по методам ИНАА и ISP содержания химических элементов Na, Ca и Cr, не значимо – Fe, Sr и Ba, а по табл.3.26 (по Ч) высоко значимо различаются по методам ИНАА и ISP содержания химических элементов Na, Ca, Fe, Sr и Ba, не значимо – Cr.

3.3.2. Двухфакторный дисперсионный анализ (п.2.3.2) соединений химических элементов в накипи разных территорий, полученных разными методами.

Рассмотрим далее применение двухфакторного дисперсионного анализа с фиксированными эффектами (п. 2.3.2) к рассматриваемому примеру (Приложение П₄). В данном случае рассматриваются факторы ОБЛ (с двумя уровнями Т и Ч) и МЕТ (с двумя уровнями ИНАА и ISP). На предварительном этапе создается в STATISTICA файл (с названием, напр., ОБЛ_МЕТ.sta) с категориальной базой данных (рис. 3.141). При этом названия факторов и переменных выносятся в названия столбцов.

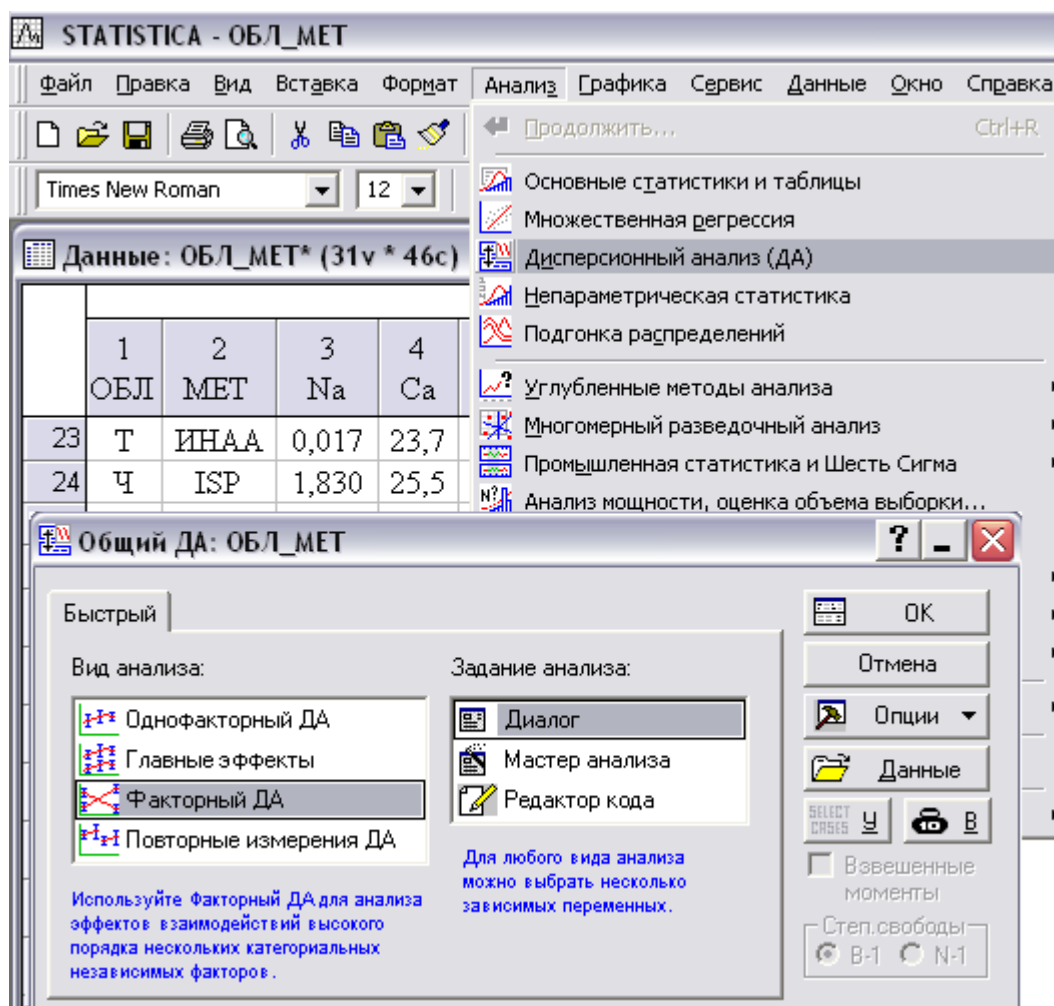
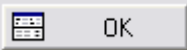
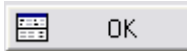


Рис.3.141. Фрагмент категориальной базы данных ОБЛ_МЕТ

Шаг 1. Запустите модуль **Анализ**, щелкните в нем по кнопке **Дисперсионный анализ (ДА)** (рис. 3.141), выберите **Факторный ДА** - тип мно-

гофакторного дисперсионного анализа, как показано на рис. 3.141, и нажмите .

Шаг 2. В появившемся окне выберите переменные (рис. 3.142: все шесть химических элементов - зависимые, ОБЛ и МЕТ – независимые с указанием всех их кодов) и нажмите .

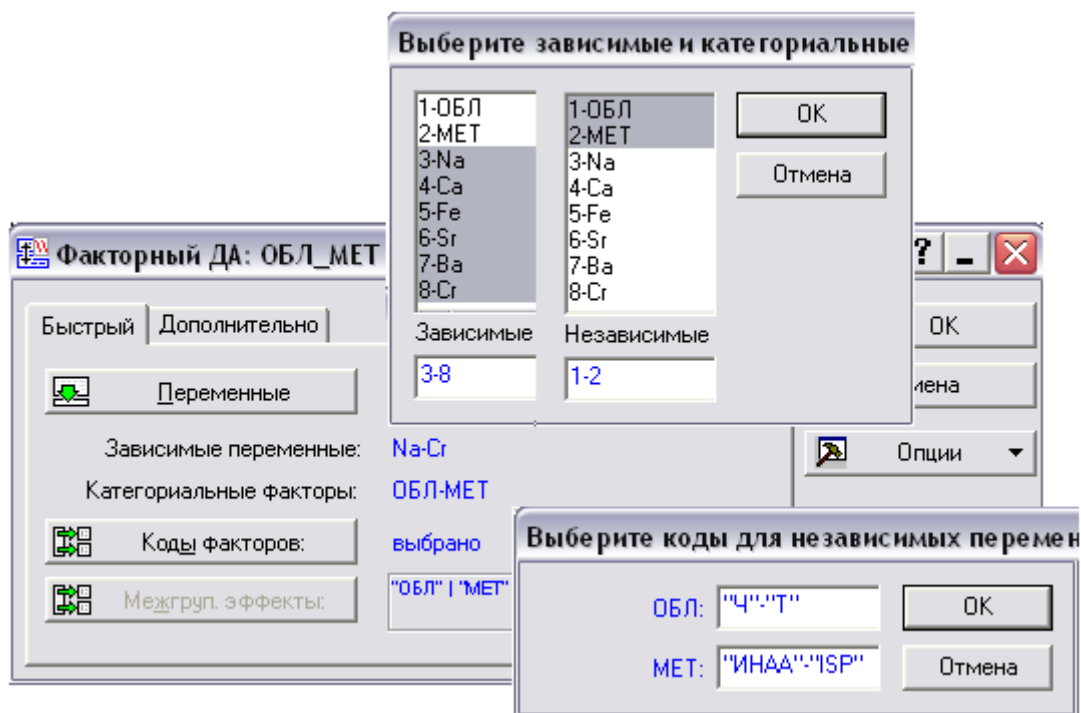

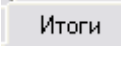
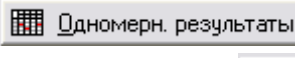



Рис. 3.142. Выбор переменных для ДА ОБЛ-МЕТ

Шаг 3. После этого в появившемся окне  (рис. 3.143) в режиме  нажмите  для просмотра результатов дисперсионного анализа (рис. 3.144) и  для просмотра результатов описательной статистики (рис. 3.145).

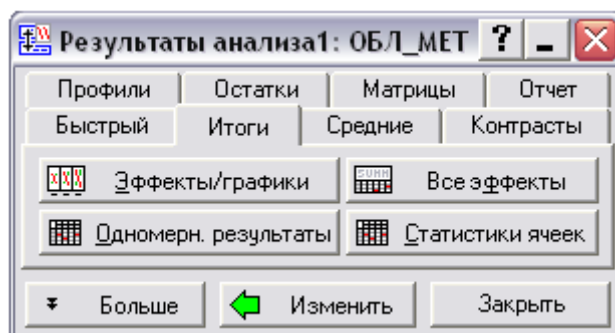


Рис. 3.143. Окно  в режиме 

Эффект	Одномерные результаты для каждой ЗП (ОБЛ_МЕТ) Сигма-ограниченная параметризация Декомпозиция гипотезы								
	Степени свободы	Ca SS	Ca MS	Ca F	Ca p	Fe SS	Fe MS	Fe F	Fe p
ОБЛ	1	569,84	569,84	5,85	0,0200	0,72	0,72	6,57	0,014
МЕТ	1	0,65	0,65	0,01	0,9352	0,48	0,48	4,36	0,043
ОБЛ*МЕТ	1	1189,14	1189,14	12,20	0,0011	0,66	0,66	6,01	0,018
Ошибка	42	4094,49	97,49			4,63	0,11		
Всего	45	5956,91				7,00			

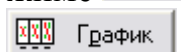
Рис. 3.144. Результаты дисперсионного анализа ОБЛ_МЕТ

Таблица на рис. 3.144 объединяет главные результаты анализа. Можно заметить, что, например, для Ca эффект фактора ОБЛ являются статистически значимым ($0,005 < p \approx 0,02 < 0,05$), эффект фактора МЕТ – не значимыми ($0,10 < p$), а эффект взаимодействия ОБЛ*МЕТ является сильно значимым ($0,0005 < p \approx 0,001 < 0,005$). В случае Fe все эффекты являются статистически значимыми ($0,005 < p < 0,05$).

Эффект	Описательные статистики (ОБЛ_МЕТ)								
	Уровень Фактор	Уровень Фактор	N	Fe Среднее	Fe Ст.откл.	Fe Стд.ош.	Fe -95,00%	Fe +95,00%	
Всего			46	0,526	0,394	0,058	0,409	0,643	
ОБЛ	Ч		30	0,434	0,347	0,063	0,304	0,564	
ОБЛ	Т		16	0,698	0,430	0,108	0,468	0,927	
МЕТ	ИНАА		23	0,671	0,368	0,077	0,512	0,831	
МЕТ	ISP		23	0,380	0,371	0,077	0,219	0,541	
ОБЛ*МЕТ	Ч	ИНАА	15	0,667	0,303	0,078	0,500	0,835	
ОБЛ*МЕТ	Ч	ISP	15	0,201	0,202	0,052	0,089	0,313	
ОБЛ*МЕТ	Т	ИНАА	8	0,679	0,493	0,174	0,267	1,091	
ОБЛ*МЕТ	Т	ISP	8	0,716	0,392	0,139	0,389	1,044	

Рис. 3.145. Результаты описательной статистики Fe

Шаг 4. Графики средних разных эффектов можно построить в режиме (рис. 3.146), выбирая нужный эффект и нажимая кнопку



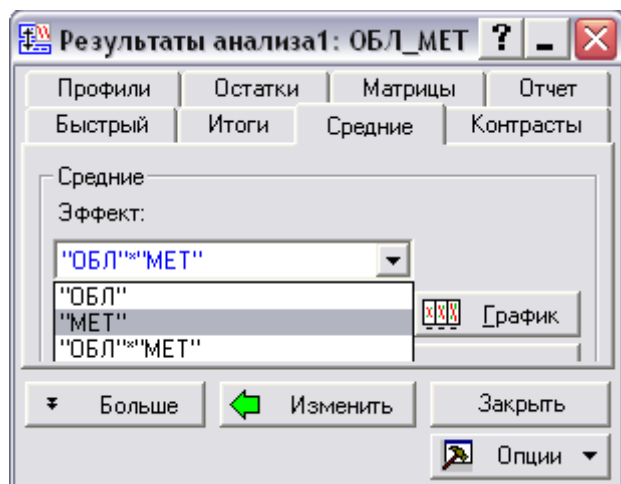


Рис. 3.146. Окно **Результаты анализа1: ОБЛ_МЕТ** в режиме **Средние**

Перебирая разные эффекты, можно построить, например, для Fe графики средних фактора МЕТ (рис.3.147), фактора ОБЛ (рис.3.148) и взаимодействия факторов МЕТ и ОБЛ (рис.3.149 или 3.150).

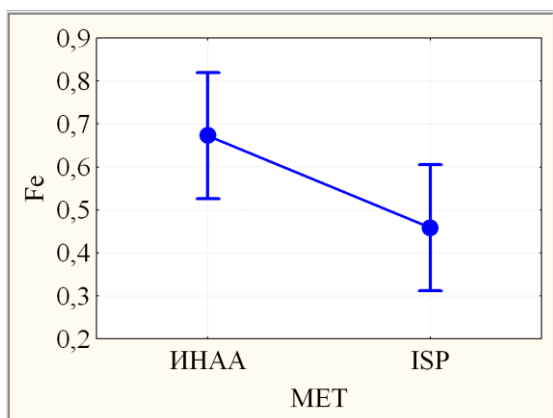


Рис. 3.147. График средних МЕТ

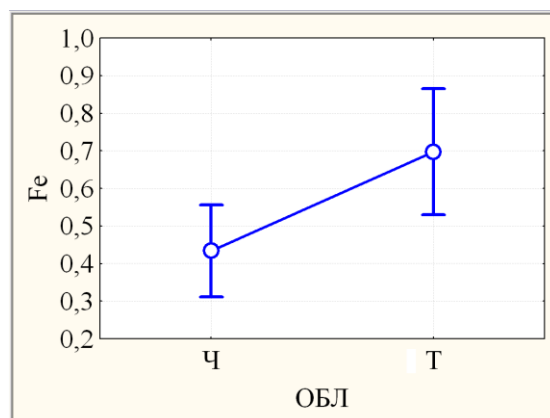


Рис. 3.148. График средних ОБЛ

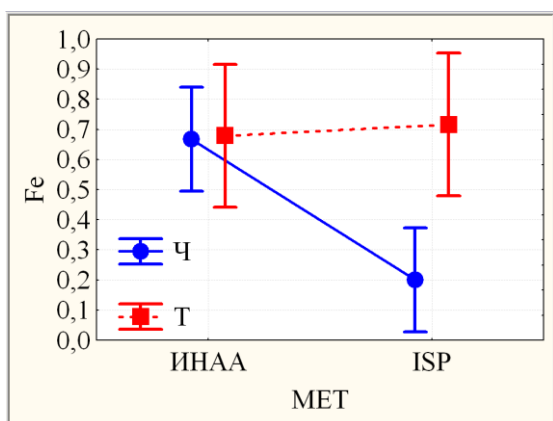


Рис. 3.149. Графики средних ОБЛ*МЕТ

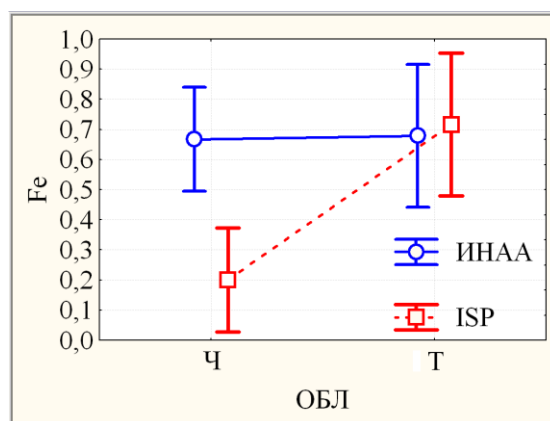

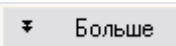
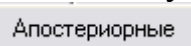
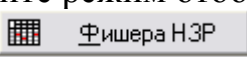


Рис. 3.150. Графики средних МЕТ*ОБЛ

Заметим, что на рис. 3.147 изображен график средних фактора МЕТ, усредненных по областям Ч и Т (см. рис.3.149), а на рис. 3.148 изображен график средних фактора ОБЛ, усредненных по методам ИНАА и ISP (см. рис.3.150). Сравнение рис. 3.147 с рис. 3.149 наглядно демонстрирует эффекты взаимодействия между факторами. Согласно рис. 3.147 среднее Fe_I меньше среднего Fe_{II} , но с учетом фактора ОБЛ выясняется, что это верно только для Ч.

Наличие статистически значимого эффекта взаимодействия предполагает далее применение метода множественных сравнений.

Шаг 5. В окне  (3.146) нажмите кнопку  и в открывшемся окне выберите режим , в котором представлены различные апостериорные критерии (рис. 3.151). Далее выберите эффект ОБЛ*МЕТ. Затем закажите режим отображения «Значимые различия» (рис. 3.152) для критерия .

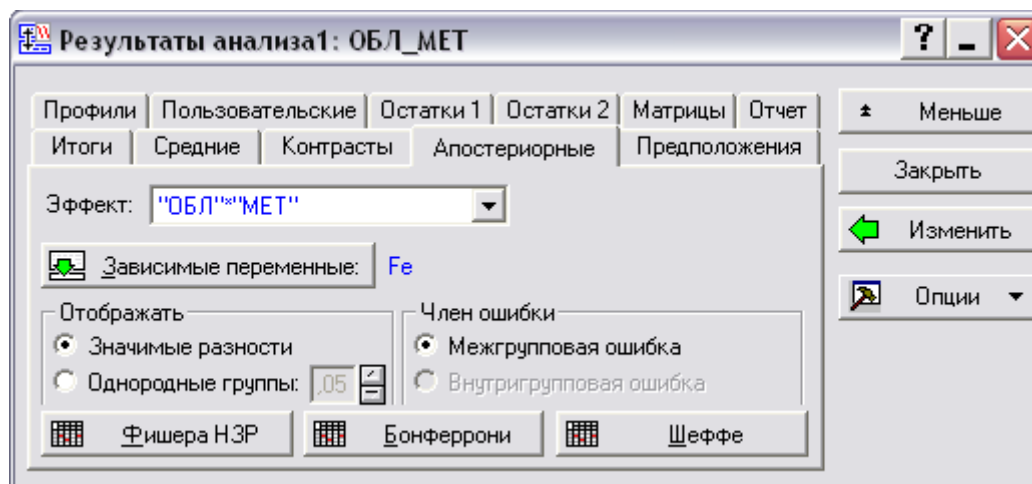
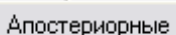


Рис. 3.151. Режим  для выбора апостериорных критериев множественных сравнений средних

N ячеек	НЗР крит.; перем. Fe (ОБЛ_МЕТ)					
	ОБЛ	МЕТ	{1}	{2}	{3}	{4}
			,66733	,20067	,67875	,71625
1	Ч	ИНАА		0,00040	0,938	0,738
2	Ч	ISP	0,00040		0,00204	0,00098
3	Т	ИНАА	0,938	0,00204		0,822
4	Т	ISP	0,738	0,00098	0,822	

Рис. 3.152. Результаты НЗР в режиме «Значимые различия»

Согласно рис. 3.152, в случае Fe для эффекта взаимодействия ОБЛ*МЕТ можно выделить однородные (различающиеся незначимо, то есть на уровне значимости $p > 0,10$) кластеры групп, расположенные в порядке убывания средних: {Т*І, Т*И, Ч*И}, {Ч*І}. Причем Ч*І отличается от остальных сильно значимо (на уровне $0,0005 < p < 0,005$). Полученные результаты находятся в полном соответствии с табл. 3.24-3.26.

Для проверки предположений, лежащих в основе параметрического дисперсионного анализа (однородность дисперсий в группах и нормальность распределении сравниваемых групп) выполните действия, аналогичные построению рис.3.73-3.75.

Шаг 6. В окне **Результаты анализа1: ОБЛ_МЕТ** выберите режим **Предположения** (рис. 3. 153), в котором для Fe выберите **Критерий Левена** (рис. 3.154) и **Гистограммы** для переменных (рис. 3.155-156).

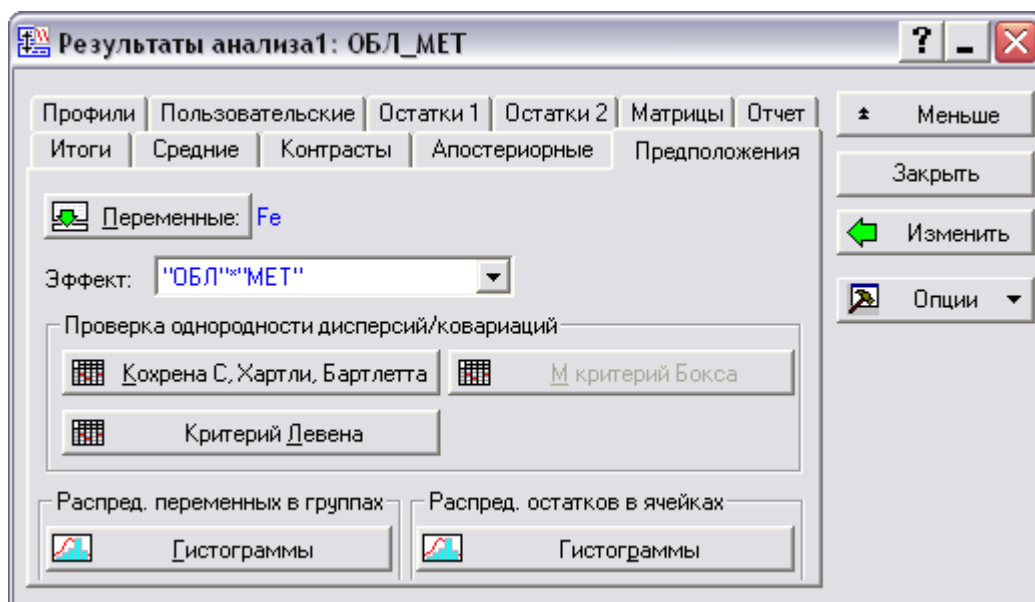


Рис. 3.153. Окно **Результаты анализа1: ОБЛ_МЕТ** в режиме **Предположения**

Критерий Левена однородности дисперсий				
Эффект: "ОБЛ*МЕТ"				
Степени свободы для всех F: 3, 42				
	MS	MS	F	
Эффект	Ошибка		p	
Fe	0,09622	0,03130	3,07404	0,03785

Рис. 3.154. Результаты проверки гипотезы об однородности дисперсий с помощью **Критерий Левена**

Согласно рис. 3.154 дисперсии различаются статистически значимо (на уровне $0,005 < p \approx 0,04 < 0,05$) по совокупности групп (за счет Fe_{ISP} по Ч и Т в табл. 3.24).

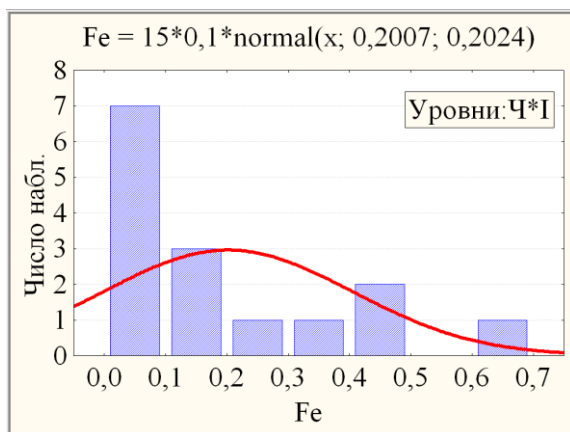


Рис. 3.155. Гистограмма $Fe_{Ч*І}$

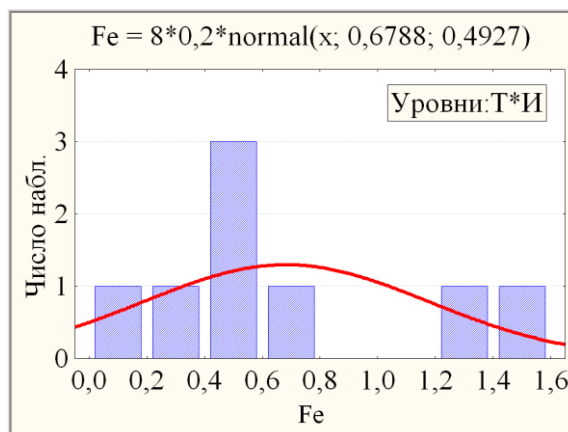


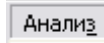
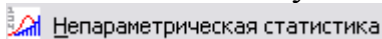
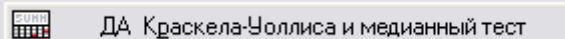
Рис. 3.156. Гистограмма $Fe_{Т*І}$

Согласно рис. 3.155-156 распределения рассматриваемых групп уровней взаимодействия факторов ОБЛ*МЕТ зримо отличаются от нормального закона.

В STATISTICA 6.1 оформлен только параметрический вариант двухфакторного дисперсионного анализа. Поэтому далее рассмотрим применение к выборке ОБЛ*МЕТ однофакторного (параметрического и непараметрического) дисперсионного анализа.

3.3.3. Однофакторный дисперсионный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.

В связи с нарушением предположений о нормальном распределении сравниваемых групп продолжим анализ данных с использованием непараметрических методов (тест Краскела-Уоллиса и медианный тест). В связи с этим построим категориальную переменную О*М из 4-х уровней взаимодействия (Ч*ИН, Ч*ИС, Т*ИН, Т*ИС) и обратимся к непараметрическим альтернативам однофакторного дисперсионного анализа (аналогично рис. 3.76 – 3.80).

Шаг 1. Запустите модуль , щелкните в нем по кнопке  и выберите множественные сравнения групп (рис. 3.157). Закажите переменные стандартным образом, беря в качестве категориальной переменной О*М, а зависимой – Fe, и щелкните по кнопке  (рис. 3.158).

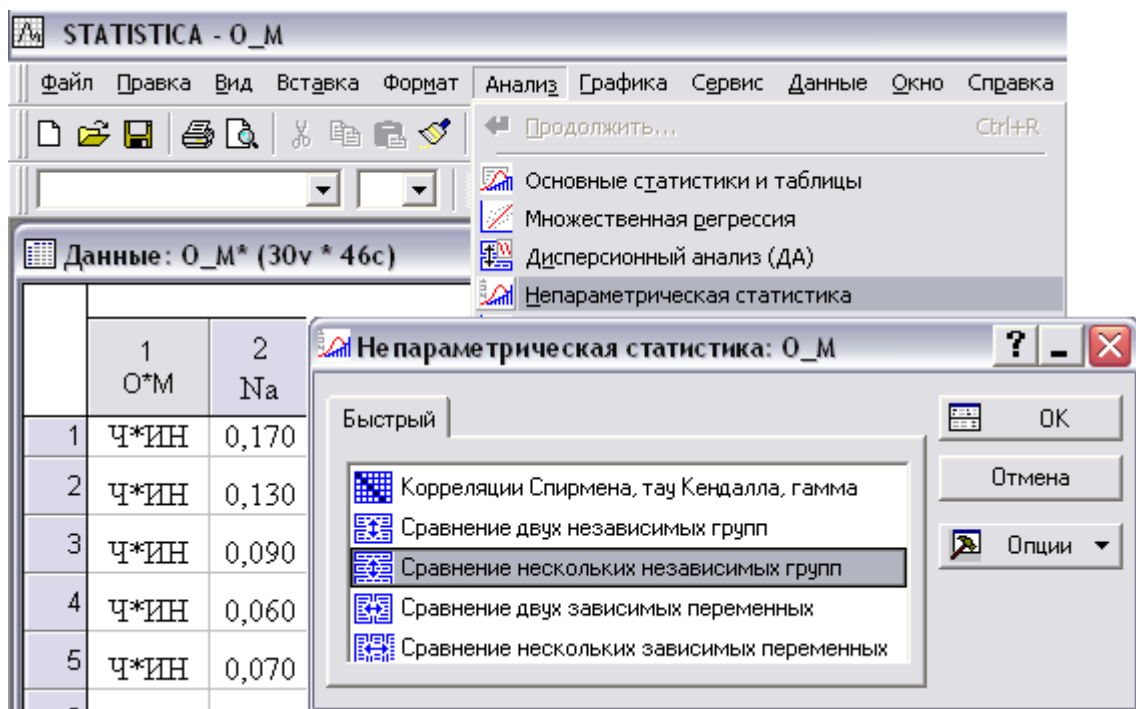


Рис. 3.157. Выбор непараметрического дисперсионного анализа O*M

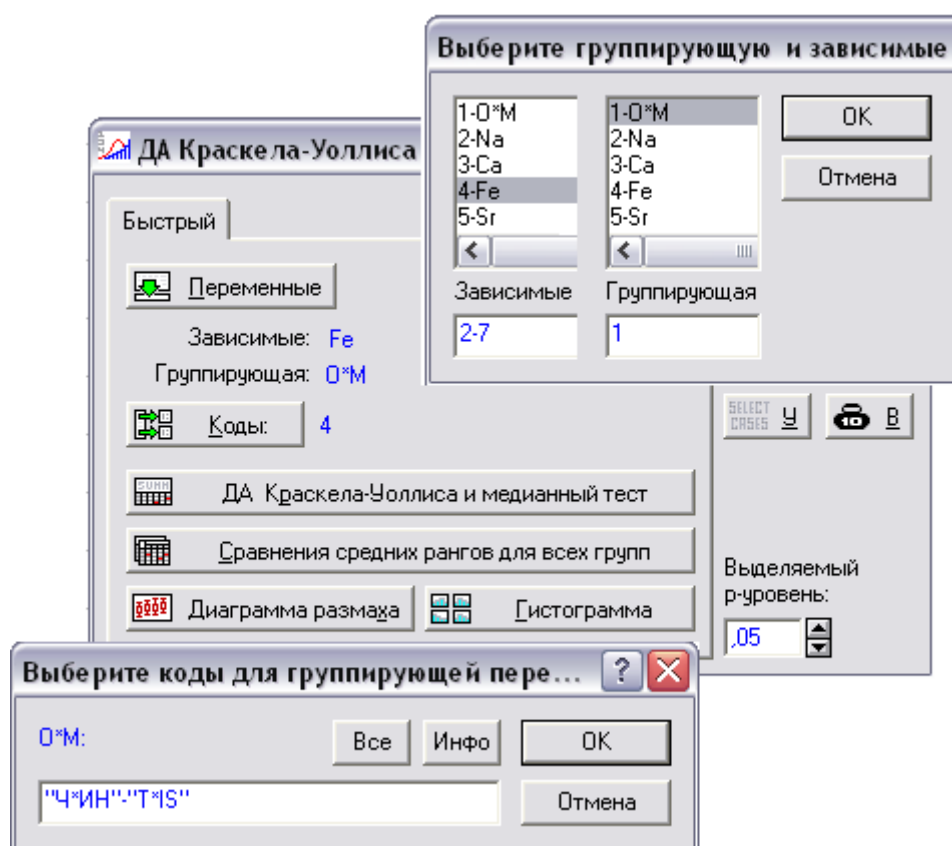



Рис. 3.158. Выбор переменных для ДА Краскела-Уоллиса

В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа. Согласно критерию Краскела-Уоллиса (рис. 3.159) существуют высоко значимые ($p \approx 0,0002 < 0,0005$), а по медианному тесту – сильно значимые ($0,0005 < p \approx 0,0007 < 0,005$), различия между средними в группах взаимодействий. Это подтверждает результат параметрического дисперсионного анализа о высоко значимой неоднородности распределения Fe по совокупности уровней фактора взаимодействия О*М и поэтому предполагает в дальнейшем использование непараметрического аналога метода множественных сравнений.

Зависим.: Fe	Ранговый ДА Краскела-Уоллиса; Fe (ОБЛ_МЕТ)		
	Группирующая переменная: О*М Кр.Краскела-Уоллиса: $H(3, N=46) = 19,52933$ $p = ,0002$		
	Код	Допуст N	Сумма Ряды
Ч*ИН	103	15	446,0000
Т*ИН	104	8	226,5000
Ч*IS	105	15	164,5000
Т*IS	106	8	244,0000

Рис. 3.159. Результаты рангового дисперсионного анализа О*М по тесту Краскела-Уоллиса

Шаг 8. В окне **ДА Краскела-Уоллиса** (рис. 3.158) нажмите кнопку

 Сравнения средних рангов для всех групп (рис. 3.160).

Зависим.: Fe	Сравнения p значений (2-стороннее); Fe (ОБЛ_МЕТ)			
	Группирующая переменная: О*М Кр.Краскела-Уоллиса: $H(3, N=46) = 19,52933$ $p = ,0002$			
	Ч*ИН R:29,733	Т*ИН R:28,313	Ч*IS R:10,967	Т*IS R:30,500
Ч*ИН		1,000	0,00077	1,000
Т*ИН	1,000		0,01896	1,000
Ч*IS	0,00077	0,01896		0,00532
Т*IS	1,000	1,000	0,00532	

Рис. 3.160. Результаты сравнений средних рангов групп О*М

Сравнивая рис. 3.160 с рис. 3.152, можно заметить, что результаты непараметрического множественных сравнений в целом выглядят более

мягкими, но тем не менее не размывающими грани однородности кластеров групп О*М. То есть и в данном случае для эффекта взаимодействия ОБЛ*МЕТ можно выделить те же однородные кластеры групп, расположенные в порядке убывания средних рангов: {Т*IS, Ч*ИН, Т*ИН} и {Ч*IS}. Причем Ч*IS отличается от Т*ИН статистически значимо (на уровне $0,005 < p \approx 0,019 < 0,050$).

Шаг 2. В окне **ДА Краскела-Уоллиса** (рис.3.158) нажмите кнопку **Диаграмма размаха** (рис. 3.158) для просмотра диаграммы размаха по группам, выберите ее тип (рис. 3.161), например, **Медиана/Квартиль/Размах** (рис. 3.162) или (для сравнения) **Среднее/Ст.Ош./Стд.Откл.** (рис. 3.163).

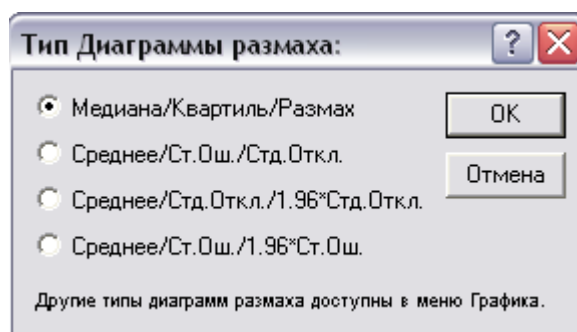


Рис. 3.161. Окно выбора типа диаграммы размаха

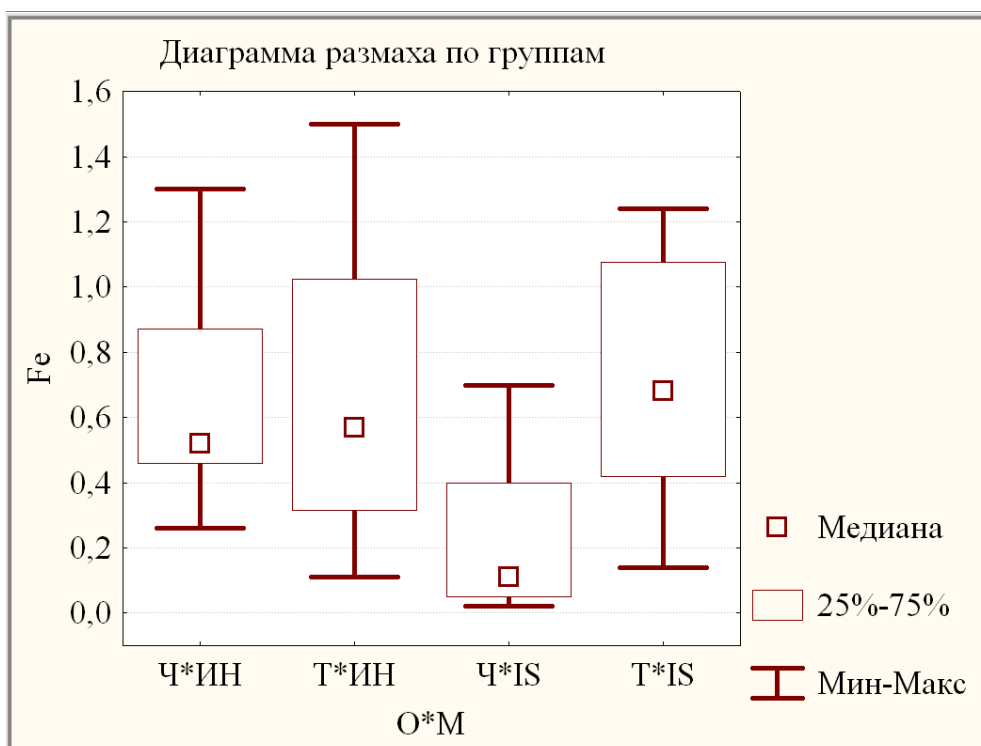


Рис. 3.162. Диаграммы размаха по группам О*М (квадрат – медиана, прямоугольник – квартильных размах, усы – абсолютный размах)

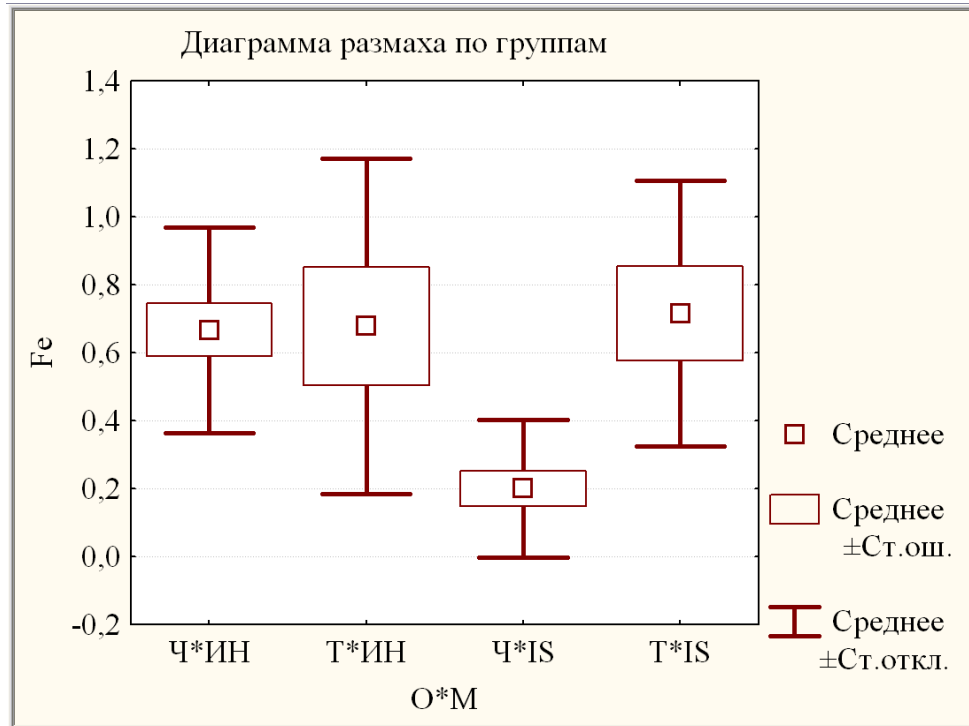


Рис. 3.163. Диаграммы размаха по группам О*М (квадрат – среднее, прямоугольник – \pm стандартная ошибка, усы – \pm стандартное отклонение)

Шаг 3. В окне **ДА Краскела-Уоллиса** (рис.3.158) нажмите кнопку



для просмотра гистограмм по группам (рис. 3.164).

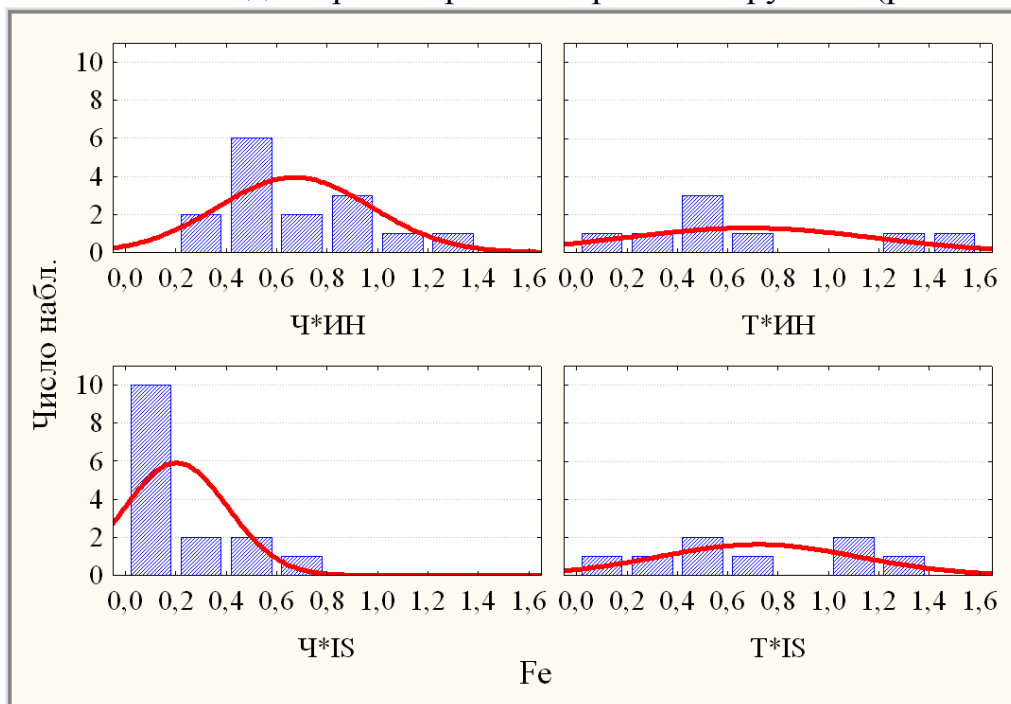


Рис. 3.164. Категориальная гистограмма распределения групп Fe

Рис. 3.164 дает наглядное представление о законах распределений групп Fe, соответствующих уровням взаимодействия, наряду с табл. 3.23, согласно которым малочисленные распределения Т*ИН, Т*IS и Ч*ИН не противоречат как нормальному, так и логнормальному законам, а распределение Ч* IS соответствует скорее логнормальному закону, чем нормальному.

Таким образом, неуверенность в распределении рассматриваемых групп по нормальному закону в силу малочисленности выборок, с одной стороны, и устойчивость параметрических критериев относительно отклонений от нормальности распределений, с другой стороны, предполагает применение как параметрических, так и непараметрических критериев. В случае их расхождения более корректными признаются результаты непараметрических критериев.

Кстати, для оценки только эффекта взаимодействия ОБЛ*МЕТ можно воспользоваться параметрическим однофакторным дисперсионным анализом применительно к категориальной переменной О*М из 4-х уровней взаимодействия (Ч*ИН, Ч*IS, Т*ИН, Т*IS) (аналогично рис. 3.64 – 3.75). В частности, вместо рис. 3.163, аналогично рис.3.70, будем иметь график средних групп Fe (рис. 3.165) с результатами НЗР (рис. 3.152), полученными аналогично рис. 3.72.

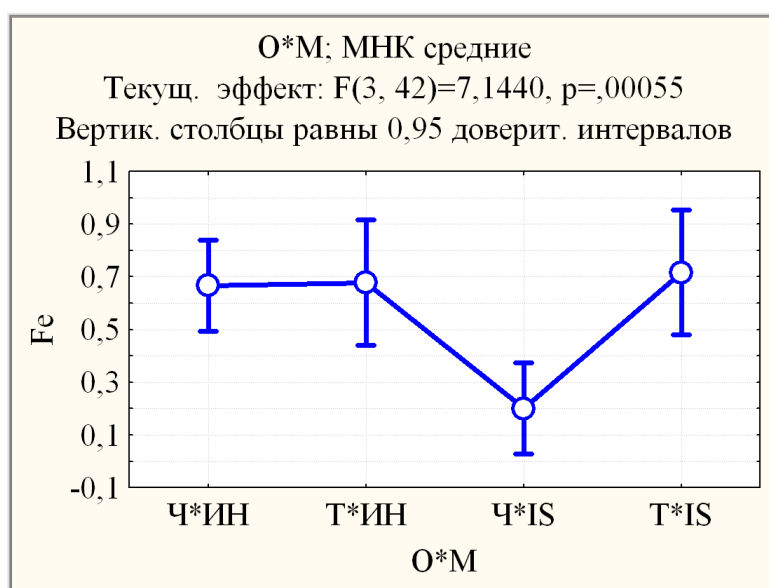


Рис. 3.165. Категориальный график средних групп Fe

Совокупные результаты однофакторного дисперсионного анализа применительно к категориальной переменной О*М на примере Са можно оформить в компактном виде рис. 3.166-3.170 аналогично соответственно рис. 3.75, 3.70, 3.80, 3.72 и 3.79:

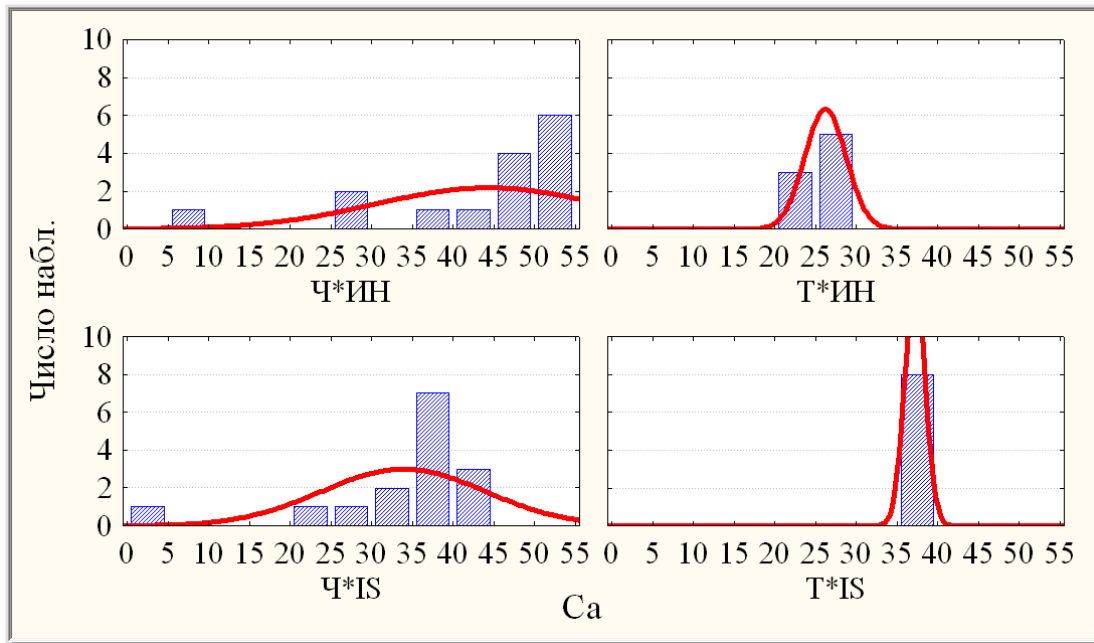


Рис. 3.166. Категориальная гистограмма распределения групп Ca

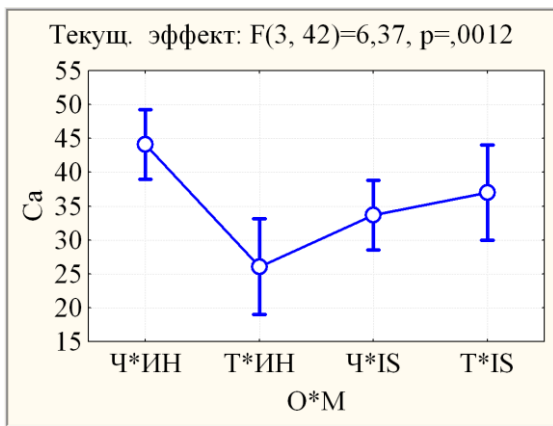


Рис. 3.167. Категориальный график средних групп Ca

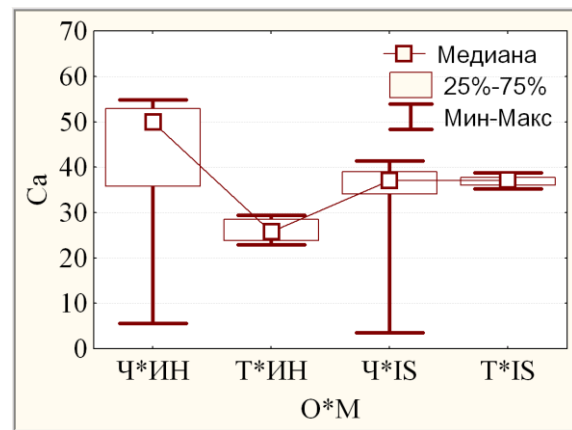


Рис. 3.168. Категориальная диаграмма размаха групп Ca

НЗР крит.; перем. Ca (ОБЛ_МЕТ) Вероятности для апостер. критериев Ошибка: Межгр. MS = 97,488, сс = 42,000				
O*M	{1}	{2}	{3}	{4}
Ч*ИН	44,140	26,075	33,715	37,000
Т*ИН		0,00014	0,00605	0,106
Ч*IS	0,00014		0,084	0,032
Т*IS	0,00605	0,084		0,451
	0,106	0,03239	0,451	

Рис. 3.169. Результаты НЗР Ca

Завис Ca	Сравнения p значений (2-сторон); Ca Группирующая переменная: p = ,0004 Кр.Краскела-Уоллиса: H(3, N= 46) = 18			
	Ч*ИН	Т*ИН	Ч*IS	Т*IS
	R:33,133	R:8,3750	R:21,700	R:23,93
Ч*ИН		0,00015	0,118	0,706
Т*ИН	0,00015		0,140	0,122
Ч*IS	0,118	0,140		1,000
Т*IS	0,706	0,122	1,000	

Рис. 3.170. Результаты сравнений средних рангов групп Ca

3.3.4. Корреляционно-регрессионный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.

Рассмотрим оценку коэффициента парной корреляции между результатами содержания химического элемента по методам ИНАА и ISP.

Корреляция между результатами содержания химического элемента по методам ИНАА и ISP характеризуется коэффициентом парной корреляции Пирсона r для нормальных распределений или ранговым коэффициентом корреляции Спирмена R , если распределения заметно отличаются от нормального. При этом оценка значимости корреляции приводит к уровню значимости α коэффициента корреляции, определяющему значимость отличия коэффициента корреляции от нуля.

Результаты вычисления коэффициентов корреляции (аналогично рис.3.84-87) между содержаниями химического элемента по методам ИНАА и ISP с учетом закона распределения содержания и оценка значимости коэффициентов корреляции сведены в табл.3.27.

Таблица 3.27

Коэффициенты корреляции между содержаниями микроэлемента в накипи по методам ИНАА и ISP с учетом закона распределения и оценка значимости коэффициентов корреляции

	Т ($n=8$)		Ч ($n=15$)	
Микро-элемент	r (α_r)	R (α_R)	r (α_r)	R (α_R)
Na		0,86 (0,006)		- 0,42 (0,12)
Ca	0,35 (0,40)			0,78 (0,001)
Fe	0,66 (0,08)		0,83 (0,000)	
Sr	0,96 (0,001)		0,93 (0,000)	
Ba	0,98 (0,000)		0,69 (0,004)	
Cr	0,02 (0,96)		0,15 (0,60)	

Таким образом, согласно табл.3.27, результаты содержаний химического элемента в накипи по методам ИНАА и ISP можно считать не

коррелированными для Na в (Ч), Ca и Fe в (Т), Cr в (Т) и (Ч); сильно коррелированными для Na в (Т), Ca и Fe в (Ч), Sr и Ba в (Т) и (Ч).

Полученные результаты наглядно иллюстрируются на составных по (Ч) и (Т) диаграммах рассеяния содержаний химических элементов в накипи по методам ИНАА и ISP посредством распыления облака значений пары (ИНАА, ISP) около прямой регрессии. Построения выполнены аналогично 3.29-3.32 или 3.91-93 для категориальной по областям (Ч) и (Т) базы данных, в которой переменными оформлены содержания химических элементов в накипи по методам ИНАА и ISP ($Na_{ин}$, $Na_{ис}$, $Fe_{ин}$, $Fe_{ис}$ и т.д.). Согласно диаграмме рассеяния, степень распыления облака значений пары (ИНАА, ISP) около прямой регрессии визуальнo характеризует силу корреляционной связи. На диаграммах рассеяния толстой сплошной замкнутой линией нарисован эллипс рассеяния и границы 75% доверительного интервала. Например, согласно табл.3.27, результаты содержаний по методам ИНАА и ISP для Na (рис.3.171) сильно коррелированными в (Т) и не коррелированными в (Ч), для Fe (рис.3.172) слабо коррелированными в (Т) и сильно коррелированными в (Ч), для Sr (рис.3.173) и Ba (рис.3.174) сильно коррелированными как в (Т), так и в (Ч).

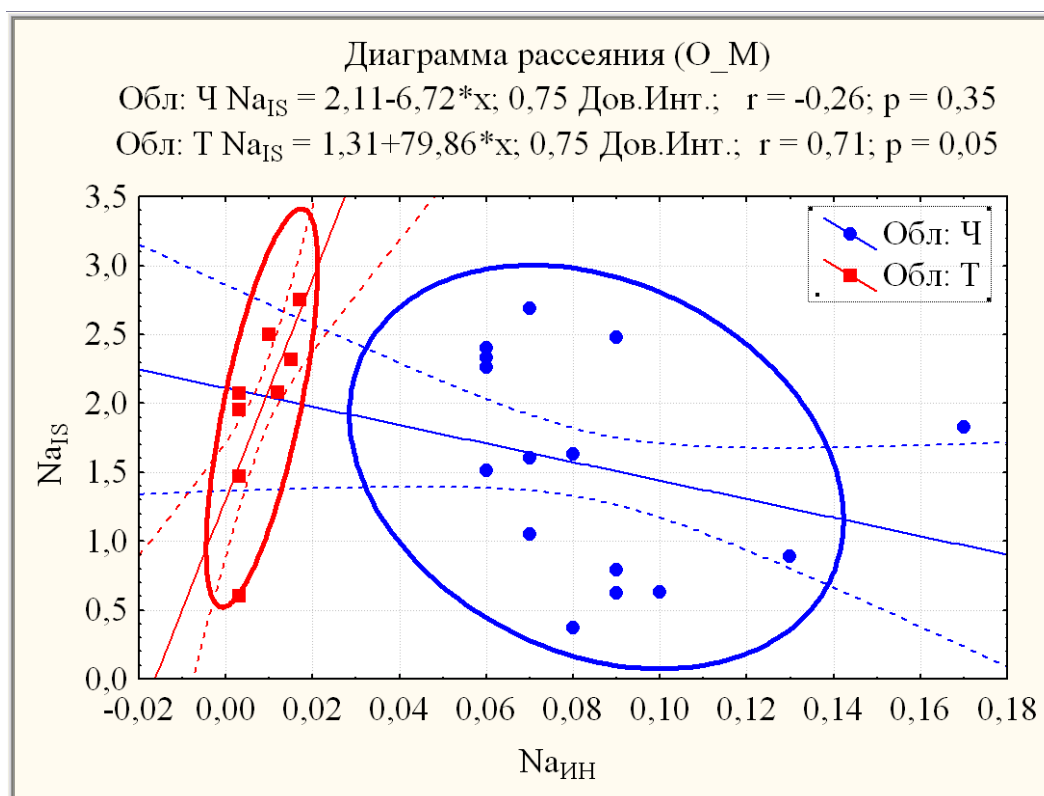


Рис. 3.171. Составная (по областям) диаграмма рассеяния содержания Na (по методам)

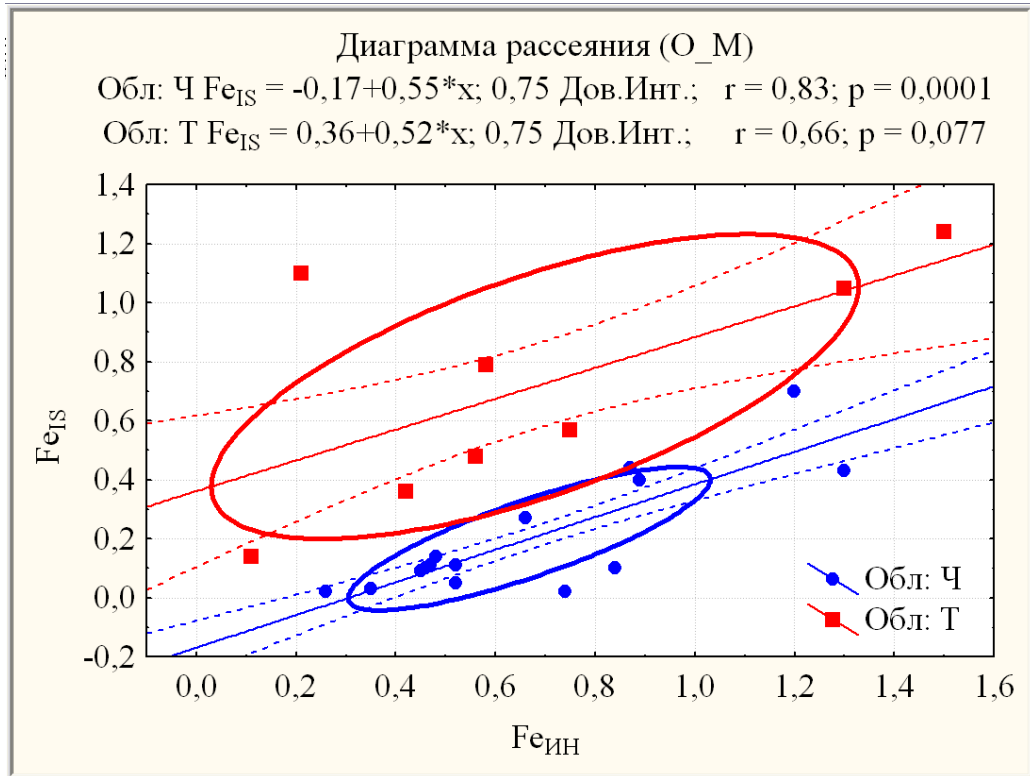


Рис. 3.172. Составная (по областям) диаграмма рассеяния содержания Fe (по методам)

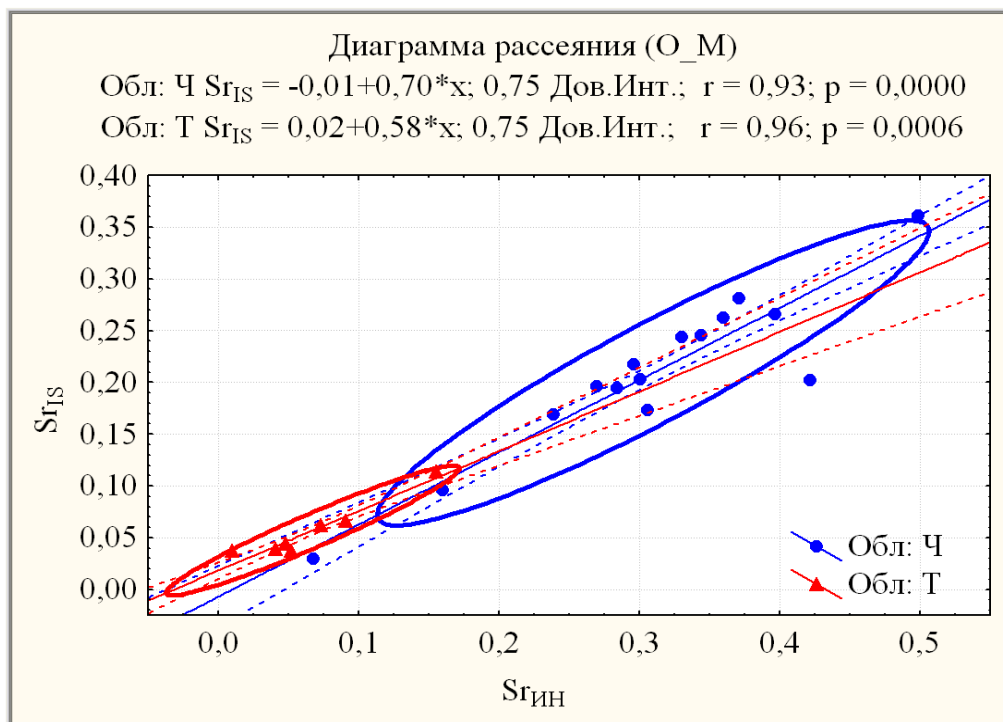


Рис. 3. 173. Составная (по областям) диаграмма рассеяния содержания Sr (по методам)

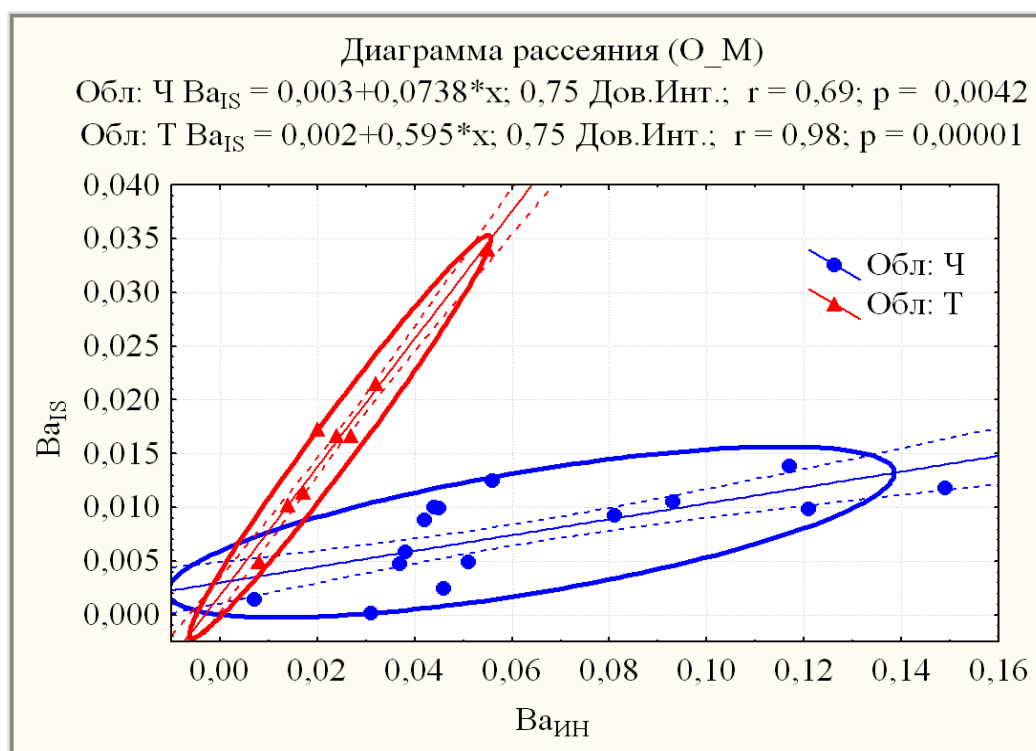


Рис. 3. 174. Составная (по областям) диаграмма рассеяния содержания Ва (по методам)

При этом составные по областям (Ч) и (Т) диаграммы рассеяния содержаний химических элементов в накипи по методам ИНАА и ISP дают также наглядное представление о степени различия в значениях средних содержаний химических элементов. Например, согласно табл.3.24-26, средние содержаний для Na (рис.3.171) различаются высоко значимо между Т и Ч по методу ИНАА, а также между методами ИНАА и ISP как в Т, так и в Ч; при этом незначимо между Т и Ч по методу ISP. Иными словами, среди содержаний Na в накипи разных территорий, полученных разными методами к однородным (различающимся незначимо) можно выборки отнести выборки Т*И и Ч*И, а к неоднородным – попарно Т*И и Ч*И, Т*И и Т*И, Ч*И и Ч*И. Аналогично для Fe (рис.3.172): Т*И и Ч*И, Т*И и Т*И – однородные, а Т*И и Ч*И, Ч*И и Ч*И – неоднородные; для Sr (рис.3.173): Т*И и Т*И – однородные, а Т*И и Ч*И, Т*И и Ч*И, Ч*И и Ч*И – неоднородные; для Ва (рис.3.174): Т*И и Т*И – однородные, а Т*И и Ч*И, Т*И и Ч*И, Ч*И и Ч*И – неоднородные.

Наряду с коэффициентами корреляции между содержаниями химического элемента по методам ИНАА и ISP (табл.3.27) можно построить корреляционную матрицу (аналогично рис. 3.84-87) ассоциации микроэлементов с учетом законов их распределения (табл.3.23) в накипи как по городам, так и по методам (табл.3.28-31).

Таблица 3.28

Корреляционная матрица ассоциации микроэлементов в накипи Ч по методу ИНАА (Ч*И) с учетом закона распределения

	Na	Ca	Fe	Sr	Ba	Cr	
Na	1,00						r
Ca	-0,11	1,00					
Fe	0,33	-0,31	1,00	0,57	0,55	0,06	
Sr	0,16	0,16		1,00	0,48	-0,28	
Ba	0,28	-0,25			1,00	-0,14	
Cr	-0,05	0,33				1,00	
	R						

Таблица 3.29

Корреляционная матрица ассоциации микроэлементов в накипи Ч по методу ISP (Ч*И) с учетом закона распределения

	Na	Ca	Fe	Sr	Ba	Cr	
Na	1,00		-0,05	-0,43	-0,17	0,95	r
Ca	-0,39	1,00					
Fe		0,00	1,00	0,63	0,64	-0,04	
Sr		0,27		1,00	0,71	-0,41	
Ba		-0,09			1,00	-0,16	
Cr		-0,46				1,00	
	R						

Таблица 3.30

Корреляционная матрица ассоциации микроэлементов в накипи Т по методу ИНАА (Т*И) с учетом закона распределения

	Na	Ca	Fe	Sr	Ba	Cr	
Na	1,00	0,02	0,29	0,60	0,48	-0,28	r
Ca		1,00	0,46	0,36	-0,29	-0,25	
Fe			1,00	0,72	0,60	-0,28	
Sr				1,00	0,73	-0,07	
Ba					1,00	-0,01	
	R						

Таблица 3.31

Корреляционная матрица ассоциации микроэлементов в накипи Т по методу ISP (Т*І) с учетом закона распределения

	Na	Ca	Fe	Sr	Ba	Cr	
Na	1,00	0,03	0,07	0,52	0,52	0,94	r
Ca		1,00	-0,53	0,12	-0,27	-0,16	
Fe			1,00	0,45	0,70	0,10	
Sr				1,00	0,85	0,50	
Ba					1,00	0,61	
	R						

Значимые коэффициенты корреляции ($r > r_{кр} = 0,51$ для $n = 15$ и $\alpha = 0,05$; $r > r_{кр} = 0,71$ для $n = 8$ и $\alpha = 0,05$) выделены жирным шрифтом: для Ч*І– пары Fe&Sr и Fe&Ba; для Ч*І – пары Fe&Sr, Fe&Ba, Sr&Ba и Na& Cr; для Т*І– пары Fe&Sr и Sr&Ba; для Т*І – пары Sr&Ba и Na&Cr.

3.1.5. Кластерный анализ содержаний химических элементов в накипи разных территорий, полученных разными методами.

В рамках кластерного анализа построим дендрограммы содержаний химических элементов в накипи, полученных разными методами отдельно по городам (геохимические спектры микроэлементов выборок Т*І, Т*І, Ч*І и Ч*І) аналогично рис.3.39-44 или 3.97-99.

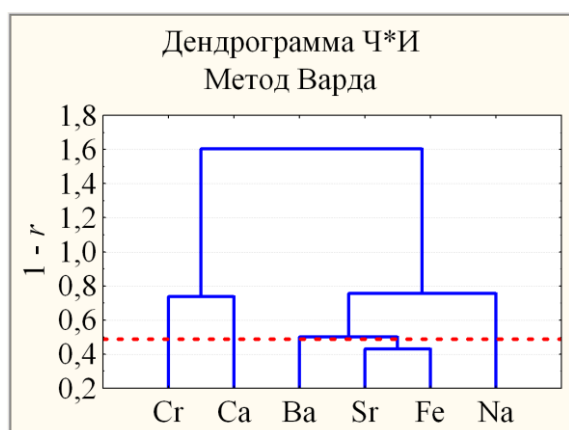


Рис. 3. 175. Дендрограмма микроэлементов в накипи Ч по методу ИНАА ($n = 15$; $1 - r_{0,05} = 0,49$)

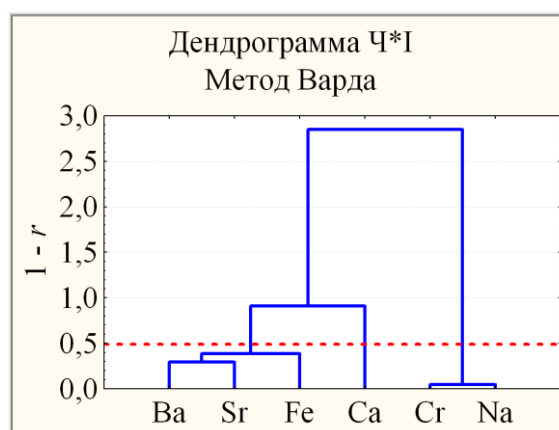


Рис. 3. 176. Дендрограмма микроэлементов в накипи Ч по методу ISP ($n = 15$; $1 - r_{0,05} = 0,49$)

На рис. 3. 175 и 3. 176 приведены дендрограммы корреляционных матриц геохимического спектра микроэлементов в накипи Ч по методам ИНАА и ISP соответственно, а на рис.3. 177 и 3. 178 – для Т. Несмотря на различия в дендрограммах, можно все же выделить одну общую по городам и методам значимую последовательную ассоциацию Ba-Sr-Fe, а так же Cr-Na по городам для метода ISP.

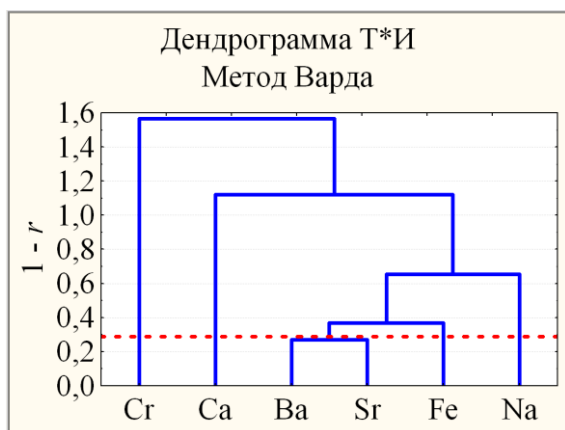


Рис. 3. 177. Дендрограмма микроэлементов в накипи Т по методу ИНАА ($1 - r_{0,05} = 0,29$; $n = 8$)

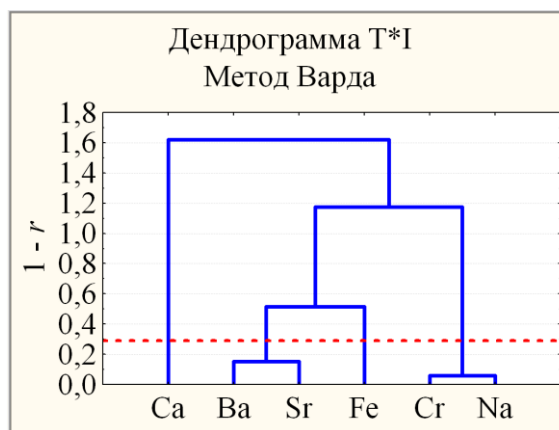




Рис. 3. 178. Дендрограмма микроэлементов в накипи Т по методу ISP ($1 - r_{0,05} = 0,29$; $n = 8$)

Заметим, что при проведении кластеризации микроэлементов в качестве меры близости микроэлементов использовано корреляционное расстояние, так как корреляционный анализ данной ассоциации микроэлементов показал наличие значимых корреляционных связей. При этом допустимы разные методы объединения кластеров. Результаты кластерного анализа микроэлементов по всем выборкам (Т*И, Т*І, Ч*И и Ч*І) демонстрируют устойчивость относительно вариации методов (Варда, полной связи, взвешенного попарного среднего) объединения кластеров.

3.3.6. Канонический анализ двух ассоциаций микроэлементов в накипи территории Ч, полученных методами ИНАА и ISP.

Применим метод канонических корреляций для анализа зависимости между ассоциациями микроэлементов выборок Ч*И и Ч*І аналогично п. 3.1.7 (рис.3.45-60).

Для анализа составной зависимости рассматриваемых переменных вызывается диалоговое окно модуля  Канонический анализ в полосе меню головной модуль  (см. рис. 3.45). Выберем базы данных выборок Ч*И и Ч*І (аналогично рис. 3.46). Среди результатов вычисления

выборочных числовых характеристик особый интерес в данном случае представляет матрицы парных коэффициентов корреляции (рис. 3.179) аналогично рис. 3.49.

	Корреляции (Ч_Мет)											
	Na _И	Ca _И	Fe _И	Sr _И	Ba _И	Cr _И	Na _Г	Ca _Г	Fe _Г	Sr _Г	Ba _Г	Cr _Г
Na _И	1,00	-0,11	0,25	0,28	0,49	0,05	-0,26	-0,07	0,31	0,32	0,26	-0,24
Ca _И	-0,11	1,00	0,07	0,51	-0,14	0,26	-0,37	0,96	-0,14	0,47	0,12	-0,44
Fe _И	0,25	0,07	1,00	0,57	0,55	0,06	-0,12	0,28	0,83	0,61	0,39	-0,20
Sr _И	0,28	0,51	0,57	1,00	0,48	-0,28	-0,46	0,62	0,49	0,93	0,64	-0,49
Ba _И	0,49	-0,14	0,55	0,48	1,00	-0,14	-0,04	0,04	0,82	0,58	0,69	-0,02
Cr _И	0,05	0,26	0,06	-0,28	-0,14	1,00	0,15	0,23	-0,14	-0,25	-0,38	0,15
Na _Г	-0,26	-0,37	-0,12	-0,46	-0,04	0,15	1,00	-0,43	-0,05	-0,43	-0,17	0,95
Ca _Г	-0,07	0,96	0,28	0,62	0,04	0,23	-0,43	1,00	0,05	0,58	0,21	-0,54
Fe _Г	0,31	-0,14	0,83	0,49	0,82	-0,14	-0,05	0,05	1,00	0,63	0,64	-0,04
Sr _Г	0,32	0,47	0,61	0,93	0,58	-0,25	-0,43	0,58	0,63	1,00	0,71	-0,41
Ba _Г	0,26	0,12	0,39	0,64	0,69	-0,38	-0,17	0,21	0,64	0,71	1,00	-0,16
Cr _Г	-0,24	-0,44	-0,20	-0,49	-0,02	0,15	0,95	-0,54	-0,04	-0,41	-0,16	1,00

Рис. 3.179. Матрица выборочных парных коэффициентов корреляции составной выборки Ч*Мет

Согласно рис. 3.179 микроассоциация Ba-Fe-Sr корреляционно значима как внутри каждой выборки Ч*И и Ч*Г, так и между ними, чего не скажешь о Na-Cr, корреляционно значимой только внутри выборки Ч*Г. Однако выделяется еще одна межгрупповая корреляционно значима микроассоциация Ca-Sr (Коэффициент парной корреляции Пирсона $r = 0,96$ между Ca_И и Ca_Г, $r = 0,93$ между Sr_И и Sr_Г, $r \approx 0,60$ между Sr_И и Ca_Г). Для визуализации корреляционной матрицы переменных микроассоциация Ca-Sr можно построить матрицу диаграмм рассеяния с гистограммами (рис. 3.180) аналогично рис. 3.50. Подобные графики представляют наглядный и эффективный способ визуального анализа тесноты зависимостей между исследуемыми переменными (степени распыления облака значений пары переменных около прямой регрессии), а гистограммы на диагонали для соответствующих переменных можно использовать для визуальной оценки закона нормального распределения переменной. На диаграммах рассеяния добавлены 75% эллипсы рассеяния (толстая сплошная замкнутая линия), границы 75% доверительного интервала (тонкая пунктирная линия) и прямые регрессии.

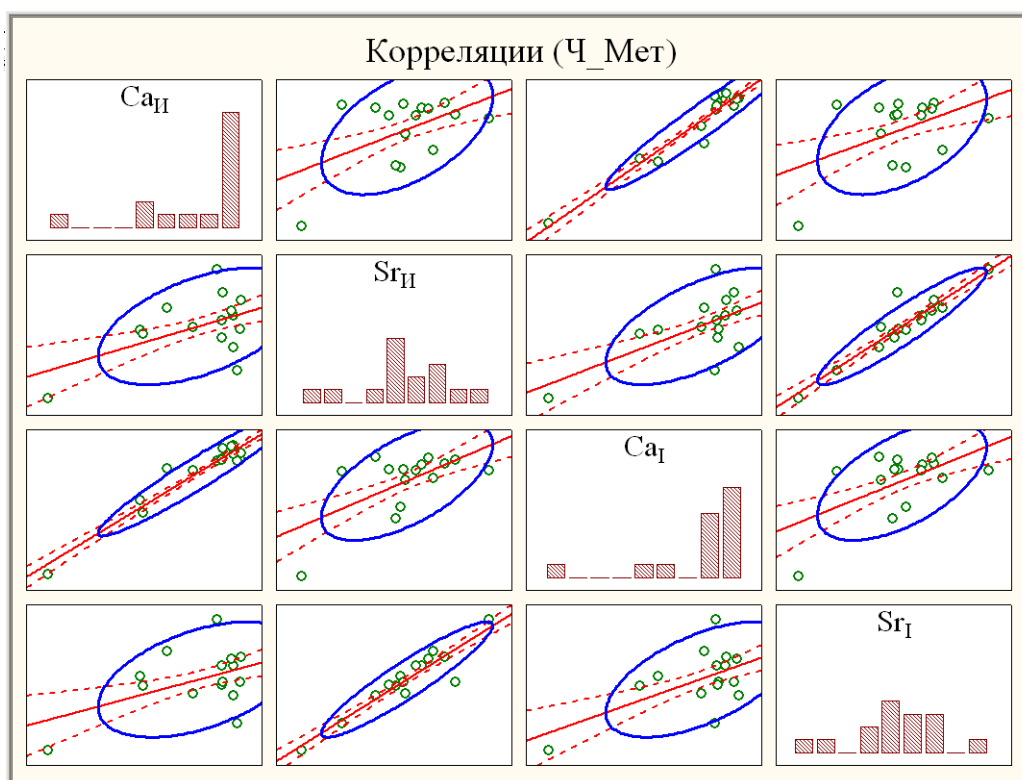


Рис. 3.180. Матрица гистограмм и диаграмм рассеяния с прямой линейной регрессии для микроассоциации Ca-Sr

Для продолжения канонического анализа выделим галочкой режим Пакетная обработка/печать и определим микроэлементы Ч*И в первую группу, а микроэлементы Ч*І - во вторую группу аналогично рис. 3.51.

Знакомство с протоколом пакетного режима выполненных вычислений начнем с Значения Хи-квадрат ... – просмотра оценок значимости канонических корреляций R_k (рис. 3. 181) аналогично рис. 3.53.

Корень Удаленный	Значения Хи-квадрат с успешно удален. корнями				
	Канонич. R	Канонич. R-кв.	Хи-кв.	сс	p
0	0,997	0,993	81,710	36	0,00002
1	0,980	0,961	44,268	25	0,01016
2	0,922	0,851	19,974	16	0,22148
3	0,677	0,458	5,705	9	0,76905
4	0,346	0,120	1,111	4	0,89256
5	0,142	0,020	0,154	1	0,69512



Рис. 3.181. Результаты оценки значимости коэффициентов канонических корреляций R_k для Ч*Мет

Согласно рис. 3.181, первому собственному значению $\lambda_1^2 \approx 0,993$ соответствует максимальный канонический коэффициент корреляции $R_1 \approx 0,997$ между первой парой U_1 и V_1 канонических переменных. В третьем столбце приводится расчетное значение статистики хи – квадрат $\chi^2 \approx 81,71$, которое при степени свободы $ss = 36$ указывает на высоко значимое ($p \approx 0,00002 < 0,0005$) отличие R_1 от нуля. Во второй строке приведена аналогичная оценка значимости коэффициента канонической корреляции $R_2 \approx 0,980$, для которого значение статистики хи – квадрат $\chi^2 \approx 44,27$ при степени свободы $ss = 25$ указывает статистически значимое ($0,005 < p \approx 0,01 < 0,05$) отличие R_2 от нуля. Эта величина может быть проинтерпретирована как корреляция между второй парой U_2 и V_2 канонических переменных.

Заметим, что последовательный критерий значимости работает в режиме последовательного удаления наиболее значимых канонических корней. Сначала рассматриваются все (6) канонические переменные вместе, т.е. без удаления корней. Полученное значение высоко значимо (на уровне значимости $p \approx 0,00002 < 0,0005$). Далее, первый (наиболее значимый) корень "удаляется" и определяется статистическая значимость 5-ти оставшихся корней. Это значение (во второй строке таблицы результатов) статистически значимо ($0,005 < p \approx 0,01 < 0,05$). Затем "удаляются" два первых значимых корней и определяется статистическая значимость 4-х оставшихся корней. Это значение (в третьей строке) незначимо ($0,10 < p \approx 0,22$). Таким образом, что только первые два корня является статистически значимым и должны быть изучены более подробно.

После вычисления собственных значений перейдем к построению соответствующих собственных векторов A_k и B_k (канонических весов). Канонические веса используются для вычисления значений канонических переменных:

$$U_k = \sum_{i=1}^6 a_{ki} X_i \quad \text{и} \quad V_k = \sum_{j=1}^6 b_{kj} Y_j.$$

В режиме «Пакетная обработка» (см. рис. 3.52) достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например,  Канон.веса, левое мн-во ... — для просмотра коэффициентов a_{ki} (рис. 3.182) аналогично рис. 3.54 и соответственно  Канон.веса, правое мн-во ... — для просмотра b_{kj} (рис. 3.183) аналогично рис. 3.54.

Переменная	Канон.веса, левое мн-во (Ч_Мет)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Na _И	-0,174	-0,166	0,433	-0,599	0,280	-0,933
Ca _И	0,908	-0,039	-0,360	-0,480	1,486	-0,441
Fe _И	-0,045	-0,732	-0,625	-1,125	0,423	0,188
Sr _И	0,098	0,031	1,080	0,773	-1,943	0,563
Ba _И	0,208	-0,240	-0,645	1,090	0,596	-0,294
Cr _И	0,125	0,381	-0,263	0,422	-1,384	-0,157

Рис. 3.182. Вектор A_k канонических весов a_{ki} для $U_k = \sum_{i=1}^6 a_{ki} X_i$

Переменная	Канон.веса, правое мн-во (Ч_Мет)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Na _Г	-0,549	-1,234	0,705	-2,098	1,857	2,374
Ca _Г	1,181	0,504	-1,076	0,402	0,031	-0,374
Fe _Г	-0,072	-0,579	-1,324	-0,107	-0,172	-0,630
Sr _Г	-0,123	-0,730	1,572	-0,620	-0,982	1,173
Ba _Г	0,114	0,157	0,096	1,108	1,080	-0,174
Cr _Г	0,736	1,401	-1,046	2,493	-2,225	-1,372

Рис. 3.183. Вектор B_k канонических весов b_{kj} для $V_k = \sum_{j=1}^6 b_{kj} Y_j$

Далее в режиме «Пакетная обработка» (см. рис. 3.52) рассмотрим обычные корреляции между каноническими переменными и переменными из каждого множества R_{UX} и R_{VY} (Факт.структ., лев.мн-во ... – рис. 3.184 аналогично рис. 3.56 и Факт.структ., прав.мн-во ... – рис. 3.185 аналогично рис. 3.57), которые называются каноническими нагрузками факторов.

Переменная	Факт.структ., лев.мн-во (Ч_Мет)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Na _И	-0,150	-0,435	0,295	-0,055	-0,112	-0,828
Ca _И	0,979	0,073	0,115	-0,139	0,048	-0,038
Fe _И	0,157	-0,871	-0,301	-0,244	-0,250	0,068
Sr _И	0,552	-0,675	0,427	0,123	-0,188	0,080
Ba _И	0,005	-0,760	-0,174	0,553	0,026	-0,293
Cr _И	0,295	0,345	-0,582	-0,170	-0,494	-0,426

Рис. 3.184. Таблица корреляций $R_{U_k X_i}$ левого множества Ч*И



Переменная	Факт.структ., прав.мн-во (Ч_Мет)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Na _I	-0,319	0,203	-0,456	0,194	-0,028	0,782
Ca _I	0,973	-0,134	0,036	-0,163	0,072	-0,056
Fe _I	-0,015	-0,905	-0,323	0,241	-0,099	-0,088
Sr _I	0,527	-0,738	0,313	0,207	-0,192	-0,022
Ba _I	0,204	-0,638	0,180	0,648	0,314	-0,011
Cr _I	-0,385	0,253	-0,405	0,364	-0,227	0,663

Рис. 3.185. Таблица корреляций $R_{V_k Y_j}$ правого множества Ч*И

Считается, что переменные, сильно коррелированные с канонической переменной, имеют с ней много общего. Поэтому канонические нагрузки факторов используются при интерпретации канонических переменных.

Согласно рис.3. 184 , с канонической переменной U_1 наиболее сильно связана положительной корреляционной связью переменная Ca_{II} ($R \approx 0,98$), а также Sr_{II} в меньшей степени ($R \approx 0,55$). Согласно рис.3. 185 , аналогично с V_1 наиболее сильно связана положительной корреляционной связью Ca_I ($R \approx 0,97$), а также Sr_I в меньшей степени ($R \approx 0,53$). В перекрестной матрице парных коэффициентов корреляции составной выборки Ч*Мет (рис. 3.179) это соответствует межгрупповой корреляционно значимой микроассоциации Ca-Sr (коэффициент парной корреляции Пирсона $r = 0,96$ между Ca_{II} и Ca_I , $r = 0,93$ между Sr_{II} и Sr_I , $r \approx 0,60$ между Sr_{II} и Ca_I).

Аналогичным образом интерпретируется вторая пара канонических переменных. Согласно рис.3. 184 , с канонической переменной U_2 наиболее сильно связана отрицательной корреляционной связью переменная Fe_{II} ($R \approx -0,87$), а также в меньшей степени Ba_{II} ($R \approx -0,76$) и Sr_{II} ($R \approx -0,68$). Согласно рис.3. 185 , аналогично с V_2 наиболее сильно связана отрицательной корреляционной связью Fe_I ($R \approx -0,91$), а также в меньшей степени Sr_I ($R \approx -0,74$) и Ba_I ($R \approx -0,64$). В перекрестной матрице парных коэффициентов корреляции составной выборки Ч*Мет (рис. 3.179) это соответствует межгрупповой корреляционно значимой микроассоциации Ba-Fe-Sr, корреляционно значимой также внутри каждой выборки Ч*И и Ч*І.

Рассмотрим, наконец, в режиме «Пакетная обработка» (см. рис. 3.52) информацию об извлеченных дисперсиях и избыточности по группе микроэлементов Ч*И ( Доли извлеч. дисперсии, лев ... — в случае левого множества) и группе микроэлементов Ч*І ( Доли извлеч. дисперсии, пра ... — в

случае правого множества) для разных корней (рис. 3.186) аналогично рис. 3.59 и 3.60.

Фактор	Доли извлеч. дисперсии левое мн-во (Ч_Мет)	
	Извлечен дисперс.	Избыточн
Корен 1	0,232910	0,231329
Корен 2	0,350858	0,337107
Корен 3	0,123694	0,105240
Корен 4	0,071880	0,032923
Корен 5	0,059668	0,007149
Корен 6	0,160990	0,003264

Переменная	Доли извлеч. дисперсии правое мн-во (Ч_Мет)	
	Извлечен дисперс.	Избыточн
Корен 1	0,252608	0,250893
Корен 2	0,315665	0,303294
Корен 3	0,101304	0,086191
Корен 4	0,119621	0,054789
Корен 5	0,033849	0,004056
Корен 6	0,176952	0,003587

Рис. 3.186. Таблица извлеченной дисперсии и соответствующей избыточности множеств Ч*И (слева) и Ч*П (справа) для разных корней

Данная информация порождает суммарные характеристики канонического анализа (п. 1.2.5):

1) *Извлеченная дисперсия.* Можно вычислить среднюю долю дисперсии, извлеченной суммой корней:

$$\frac{1}{6} \sum_{k=1}^6 \sum_{i=1}^6 R_{U_k X_i}^2 = 100\% \text{ для левого мн-ва и}$$

$$\frac{1}{6} \sum_{k=1}^6 \sum_{j=1}^6 R_{V_k Y_j}^2 = 100\% \text{ для правого мн-ва.}$$

Заклучение о суммарной по корням доле объясняемой дисперсии можно сделать, рассматривая нагрузки канонических факторов, то есть корреляции R_{UX} и R_{VY} между каноническими переменными и исходными переменными в соответствующем множестве. Если возвести эти корреляции в квадрат, то полученные числа будут отражать долю дисперсии, объясняемую каждой переменной в случае каждого корня. Для каждого корня можно вычислить среднее значение этих долей (первые столбцы в таблицах рис. 3.186) и просуммировать по всем корням. При этом получится суммарная (по корням) средняя доля изменчивости объясненной в этом множестве на основании соответствующих канонических переменных. Результат очевидный, так как все 6 корней извлекают 100% дисперсии как из левого множества переменных, так и из правого, поскольку программа извлекает в общем случае ровно столько корней,

сколько переменных в меньшем множестве, то есть 9 в нашем случае. Если количества переменных в множествах разное, то только для одного из двух рассматриваемых множеств (с меньшим числом переменных) имеется столько канонических переменных, сколько исходных переменных в нем содержится, что гарантирует 100% результат извлеченной дисперсии в этом множестве.

2) *Избыточность*. Коэффициент канонической корреляции R_k при возведении в квадрат дает долю дисперсии, общей для двух множеств. Если умножить эту долю на долю извлеченной дисперсии, то можно получить меру избыточности множества переменных, т.е., величину, показывающую, насколько избыточно одно множество переменных, если задано другое множество. Отметим, что можно вычислить избыточность первого (левого) множества переменных при заданном втором (правом) множестве, и избыточность второго (правого) множества переменных при заданном первом (левом) множестве. Поскольку последовательно извлекаемые канонические корни не коррелированы между собой, то можно просто просуммировать избыточности по всем (или только по значимым) корням, получив при этом общий коэффициент избыточности:

$$\frac{1}{6} \sum_{k=1}^6 \sum_{i=1}^6 R_{U_k X_i}^2 R_k^2 \approx 71,70 \% \quad \text{для левого мн-ва и}$$

$$\frac{1}{6} \sum_{k=1}^6 \sum_{j=1}^6 R_{V_k Y_j}^2 R_k^2 \approx 70,28 \% \quad \text{для правого мн-ва.}$$

Полученные значения избыточности можно объяснить следующим образом: пользуясь значениями всех канонических корней и получив значения переменных в правом множестве, можно объяснить, в среднем, 71,70 % дисперсии переменных в левом множестве. Аналогично, можно объяснить 70,28 % изменчивости в правой группе по значениям переменных в левом множестве. Эти результаты говорят о достаточно сильной зависимости между ассоциациями микроэлементов двух множеств Ч*И и Ч*І.

3.2.7. Факторный анализ содержаний микроэлементов в наипи территории Т, полученных методом ISP (выборка Т*І).

Главными целями факторного анализа являются сокращение числа рассматриваемых микроэлементов и определение структуры взаимосвязей между микроэлементами, т.е. классификация микроэлементов.

С использованием пакета STATISTICA 6.1 данная задача может быть решена в следующей последовательности.

Шаг 1. Запустите **Анализ**, нажмите **Многомерный разведочный анализ**, где далее выберите **Факторный анализ** аналогично рис. 3.101.

Шаг 2. Выберите **Переменные:** (рис. 3.187) отдельной базы данных T*I в открывшемся окне **Факторный анализ**. В результате появится рис. 3. 188.

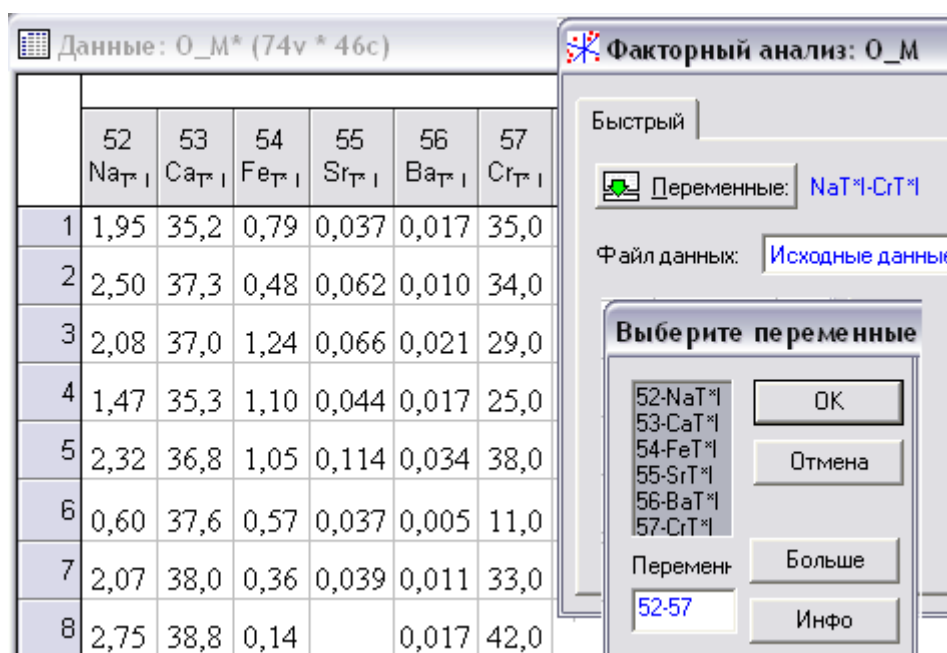


Рис. 3.187. Выбор переменных выборка T*I для факторного анализа

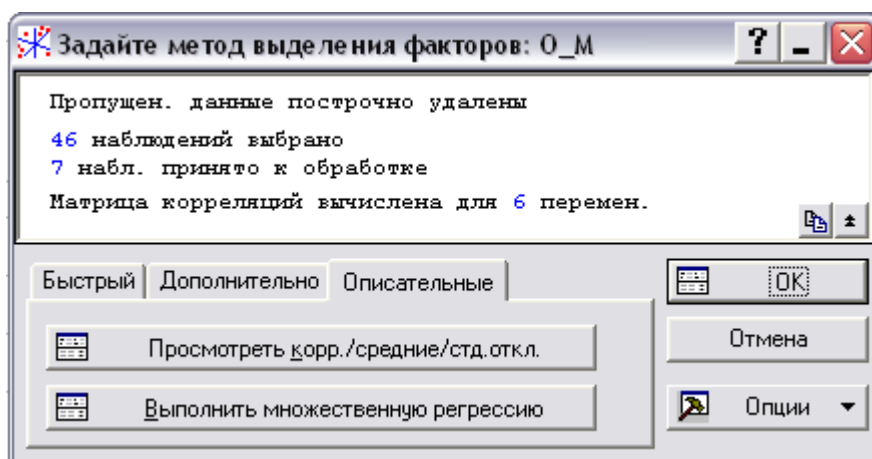

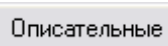


Рис. 3.188. Выбор режима **Описательные** для просмотра описательной статистики T*I

Шаг 3. Нажмите кнопку  Просмотреть корр./средние/стд.откл. в режиме  Описательные (рис. 3.188). В результате появится рис. 3.189.

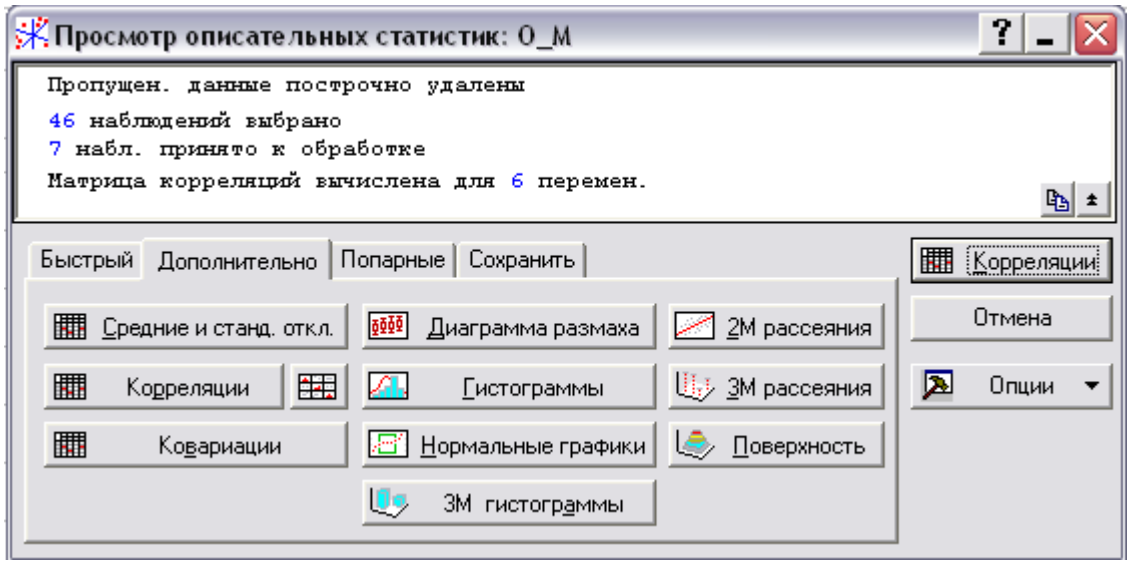
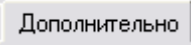
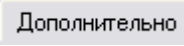
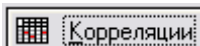
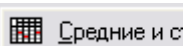
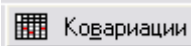
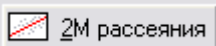
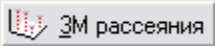


Рис.3.189. Окно для просмотра описательной статистики T*I в режиме

 Дополнительно

Можно посмотреть (рис.3.189) в режиме  Дополнительно уже знакомую корреляционную матрицу  Корреляции (см. табл.3.31) и прочую описательную статистику выборки ( Средние и станд. откл. и  Ковариации), построить различные графики, например,  2М рассеяния Na – Cr (рис. 3.190) или  3М рассеяния Ba – Sr – Fe (рис. 3.191).

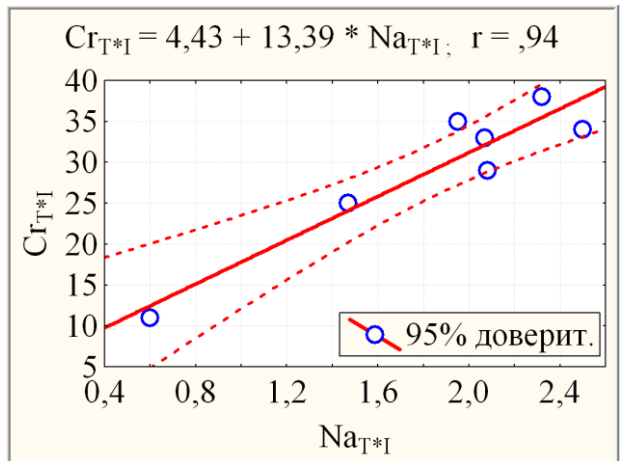


Рис.3.190. 2М Диаграмма рассеяния Na_{T*I} – Cr_{T*I}

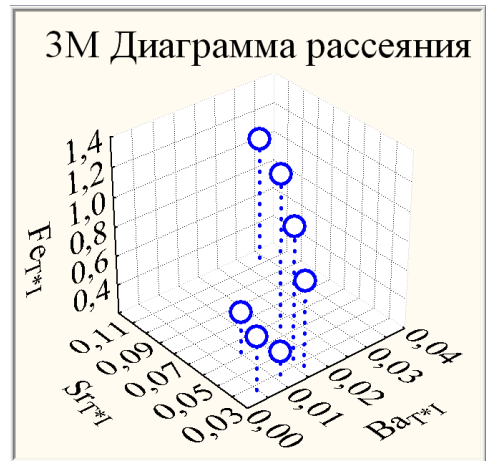


Рис.3.191. 3М Диаграмма рассеяния Ba_{T*I} – Sr_{T*I} – Fe_{T*I}

Шаг 4. Описательную статистику можно посмотреть также в режиме **Сохранить** (см. рис. 3.107). Нажмите **Матрицу** для вывода на экран корреляционной матрицы, расширенной строками средних и стандартных отклонений (рис. 3.192).

	O_M					
	1	2	3	4	5	6
	Na_T*I	Ca_T*I	Fe_T*I	Sr_T*I	Ba_T*I	Cr_T*I
Na_T*I	1,000	0,027	0,066	0,524	0,515	0,942
Ca_T*I	0,027	1,000	-0,530	0,117	-0,267	-0,161
Fe_T*I	0,066	-0,530	1,000	0,447	0,704	0,095
Sr_T*I	0,524	0,117	0,447	1,000	0,847	0,499
Ba_T*I	0,515	-0,267	0,704	0,847	1,000	0,612
Cr_T*I	0,942	-0,161	0,095	0,499	0,612	1,000
Средние	1,856	36,746	0,799	0,057	0,016	29,286
Стд. откл	0,640	1,087	0,340	0,028	0,009	9,105
Кол-во N	7,000					

Рис. 3.192. Наблюдаемая корреляционная матрица, расширенная строками средних и стандартных отклонений выборки T*I

Шаг 5. Для проведения факторного анализа выборки вернитесь в окно **Задайте метод выделения факторов:** (рис. 3.188) с помощью кнопки **Отмена** и выберите режим **Дополнительно** (рис. 3.193), в котором в поле справа можно указать Максимальное число факторов, которые будут выделены системой, например, в начале исследования 6, и Минимальное собственное значение, например 0,00 (рис. 3.193).

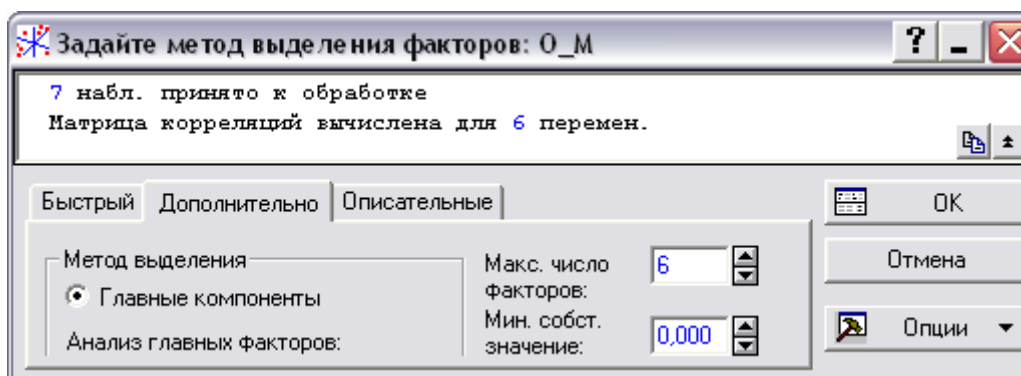


Рис. 3.193. Режим **Дополнительно** для выбора метода анализа

Собственные значения меньше указанного игнорируются. В окне **Метод выделения** выберите в качестве метода анализа **Главные компоненты** и щелкните **OK**. На экране появятся результаты анализа (рис. 3.194).

В верхней части окна **Результаты факторного анализа** дается информационное сообщение, в частности, приводятся значения собственных значений: 3,21964; 1,56157; 0,944816; 0,191459 ...

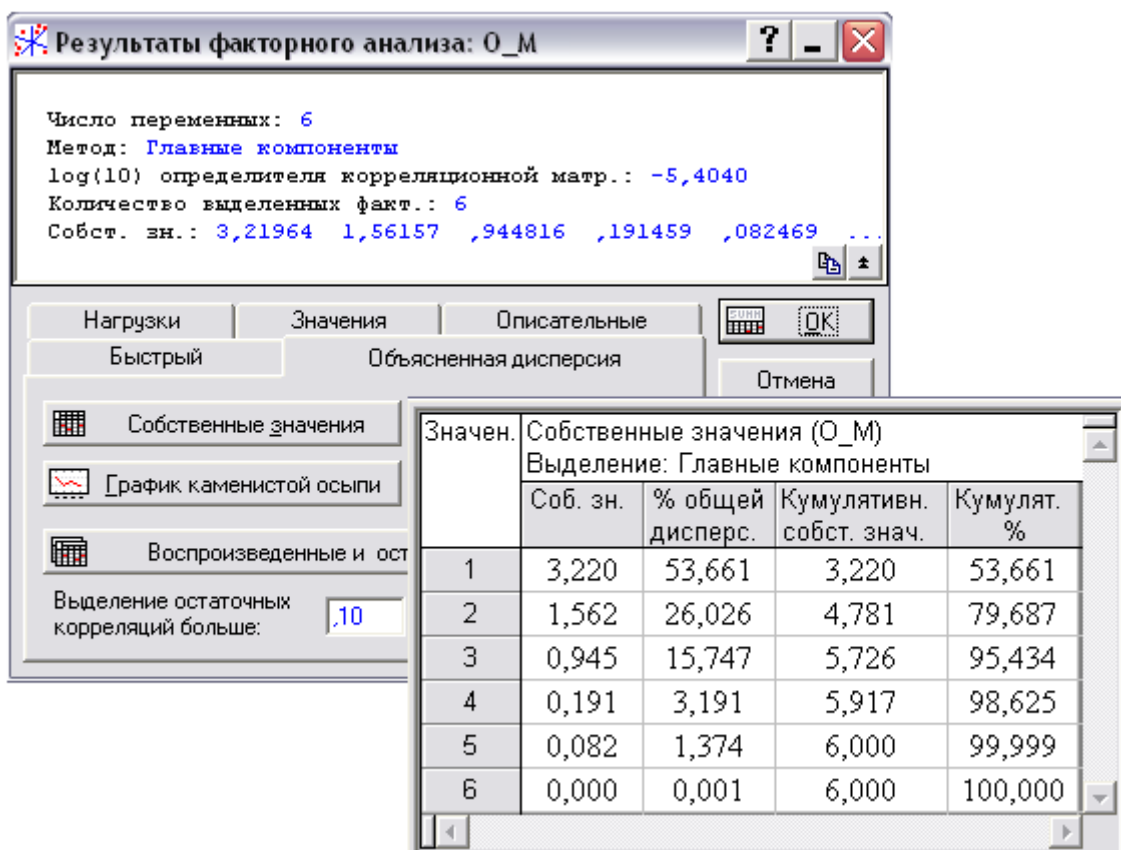


Рис. 3.194. Результаты факторного анализа для выборки T*I

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

Шаг 6. Например, для просмотра расширенной информации о собственных значениях в режиме **Объясненная дисперсия** (рис.3.194) щелкните по кнопке **Собственные значения**. Дисперсия, объясненная последовательностью собственных значений, предстанет в форме рис. 3.194.

В первом столбце (рис. 3.194) приведены собственные значения, во втором - % от общей дисперсии для каждого собственного значения, в третьем – накопленные собственные значения и в четвертом – их % от общей дисперсии. Таким образом, первое собственное значение объяс-

няет $\approx 53,66\%$ от общей дисперсии, совокупность первых двух собственных значений – уже $\approx 79,69\%$, а совокупность первых трех собственных значений объясняет $\approx 95,43\%$ от общей дисперсии.

Для проведения графического «теста каменной осыпи» выполните следующий шаг:

Шаг 7. В режиме (рис. 3.194) щелкните по кнопке для просмотра графической информации о всех собственных значениях (рис. 3.195).

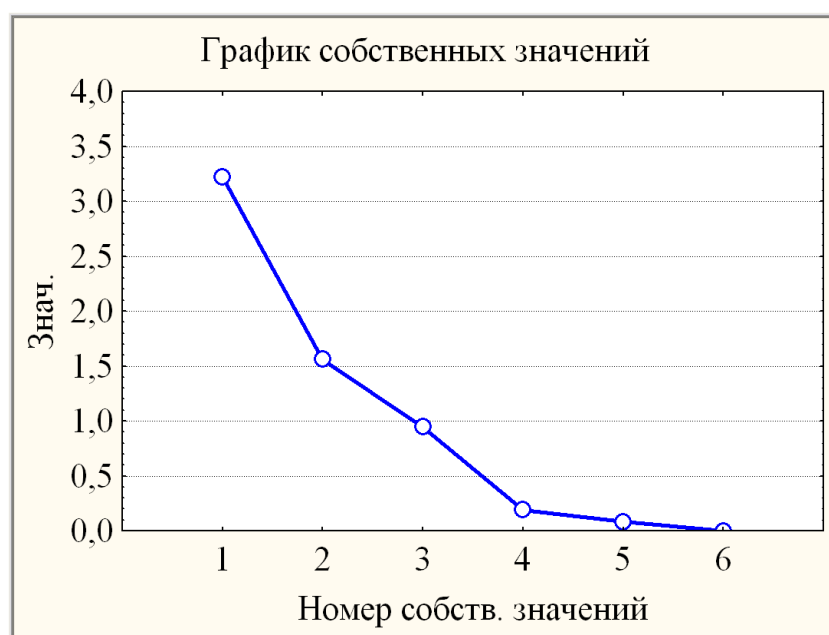
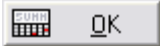
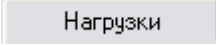


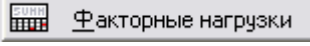


Рис. 3.195. График собственных значений

Рис. 3.195 демонстрирует графический метод («критерий каменной осыпи») оценки значимости собственных значений: справа от критической точки, где убывание собственных значений слева направо максимально замедляется, находится только «факториальная осыпь», которой пренебрегают. В соответствии с этим критерием можно оставить 3 фактора.

Самую важную информацию, на которой базируется интерпретация факторов, содержат факторные нагрузки, которые могут интерпретироваться как корреляции между факторами и переменными. Посмотрим на вращаемые факторные нагрузки в 3-х факторной модели:

Шаг 8. Щелкните кнопкой , чтобы возвратиться в окно (рис. 3.193) в режиме , в котором в поле справа укажите Максимальное число факторов, равное 3

и щелкните . На экране появятся результаты анализа 3-х факторной модели (рис. 3.196) в режиме , где выберите   и щелкните по кнопке  для просмотра факторных нагрузок в выбранной повернутой 3-х факторной модели (рис. 3.196).

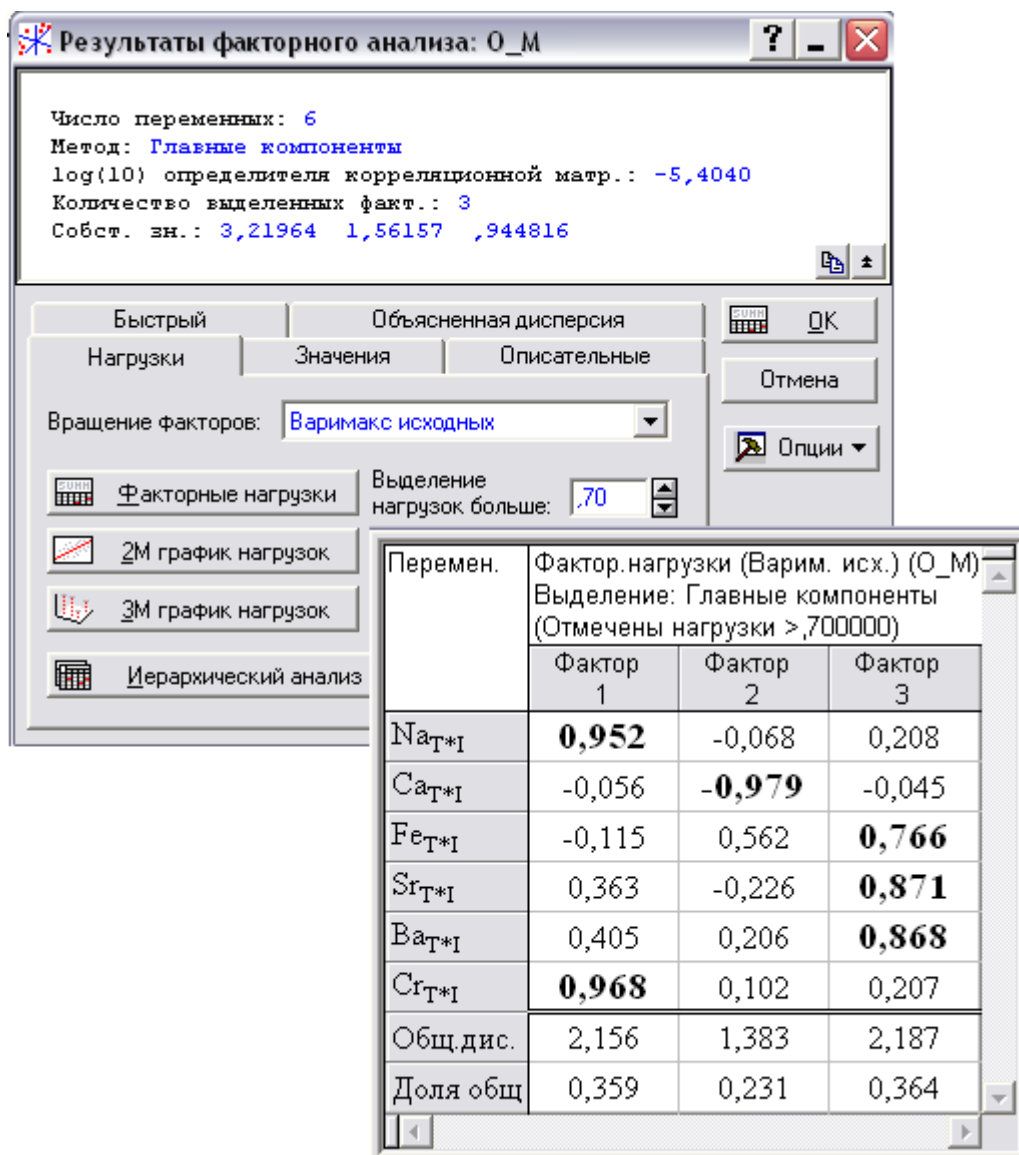


Рис. 3.196. Режим  в окне 

Для предания факторам простой интерпретации пространство факторов подвергнуто стандартному вычислительному методу варимакс


- вращения так, что факторы отмечены высокими нагрузками для одних переменных и низкими нагрузками для других.

Согласно рис. 3.196, расположенные в порядке убывания высокие факторные нагрузки микроэлементов распределились по факторам следующим образом:

Фактор Ф3 – наиболее весомый, составной, характеризуется прежде всего микроассоциацией Ba – Sr, а также микроэлементом Fe, связанных с Ф3 положительной корреляционной связью.

Фактор Ф1 – менее весомый, составной, характеризуется микроассоциацией Na – Cr, связанной с Ф1 положительной корреляционной связью.

Фактор Ф2 – еще менее весомый, характеризуется в основном микроэлементом Ca, связанным с Ф2 отрицательной корреляционной связью.

Шаг 9. Для геометрической интерпретации факторной модели щелкните по кнопке  ЭМ график нагрузок (рис. 3.197).

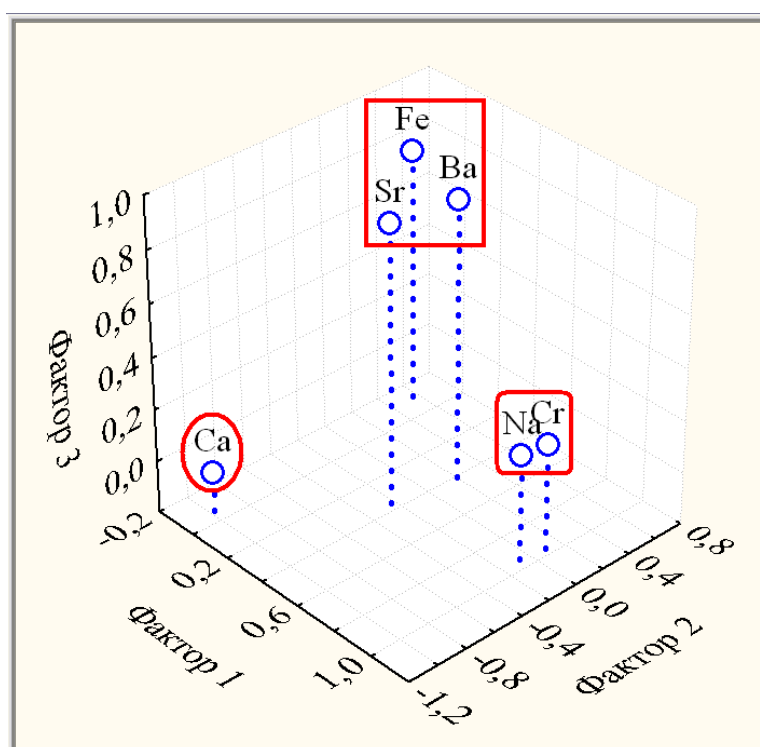

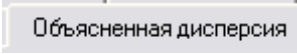
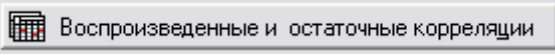


Рис. 3.197. Диаграмма рассеяния факторных нагрузок микроэлементов выборки Т*І в координатах {Ф1, Ф2, Ф3}

На рис. 3.197 изображена 3-х мерная диаграмма рассеяния факторных нагрузок микроэлементов выборки Т*І в координатах {Ф1, Ф2,

Ф3}, достаточно наглядно разделяющая 6 микроэлементов на 3 микроассоциации.

Для проведения дополнительного способа проверки достаточного числа выделенных факторов посмотрите воспроизведенную корреляционную матрицу, которая должна быть близка исходной.

Шаг 10. В полученном ранее окне  В уже знакомом режиме  (рис. 3.194) щелкните по кнопке  (рис. 3.198)

Перемен.	Воспроизведенные корреляции (O_M)					
	Na _{T*I}	Ca _{T*I}	Fe _{T*I}	Sr _{T*I}	Ba _{T*I}	Cr _{T*I}
Na _{T*I}	0,95	0,00	0,01	0,54	0,55	0,96
Ca _{T*I}	0,00	0,96	-0,58	0,16	-0,26	-0,16
Fe _{T*I}	0,01	-0,58	0,92	0,50	0,73	0,10
Sr _{T*I}	0,54	0,16	0,50	0,94	0,86	0,51
Ba _{T*I}	0,55	-0,26	0,73	0,86	0,96	0,59
Cr _{T*I}	0,96	-0,16	0,10	0,51	0,59	0,99

Рис. 3.198. Воспроизведенная корреляционная матрица выборки T*I

Сравнение исходной наблюдаемой (рис. 3.192) и воспроизведенной (рис. 3.198) корреляционных матриц выборки T*I на примере пары Fe – Sr можно оформить аналогично рис. 3.117:

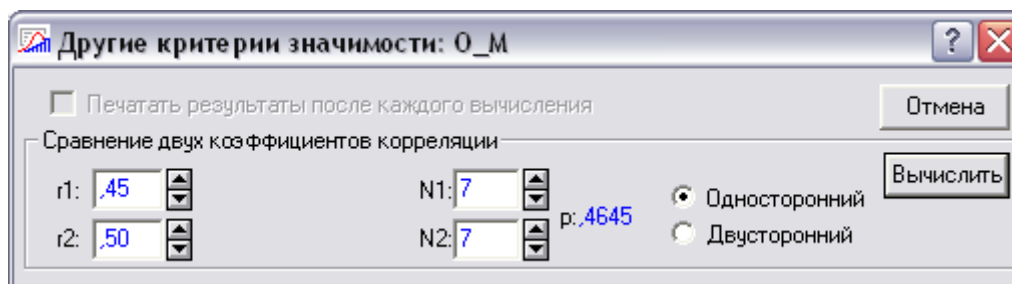


Рис. 3.199. Сравнение исходного и воспроизведенного коэффициентов корреляции пары Fe – Sr

Уровень значимости $p \approx 0,46 > 0,10$ соответствует незначимому отличию. Таким образом, 3-х факторная модель обеспечивает воспроизведение исходных переменных с незначимыми отличиями. Заметим, что согласно рис. 3.194 первые 3 собственные значения объясняют $\approx 95,43\%$ от общей дисперсии.

Для проведения дальнейшего статистического анализа выборки T*I в рамках построенной 3-х факторной модели следует сохранить значения наблюдений в новой системе факторных координат.

Шаг 11. В окне **Результаты факторного анализа** войдите в режим **Значения** (рис. 3.200) и щелкните по кнопке **Сохранить значения** для сохранения значений факторов по каждому наблюдению (рис. 3.200)

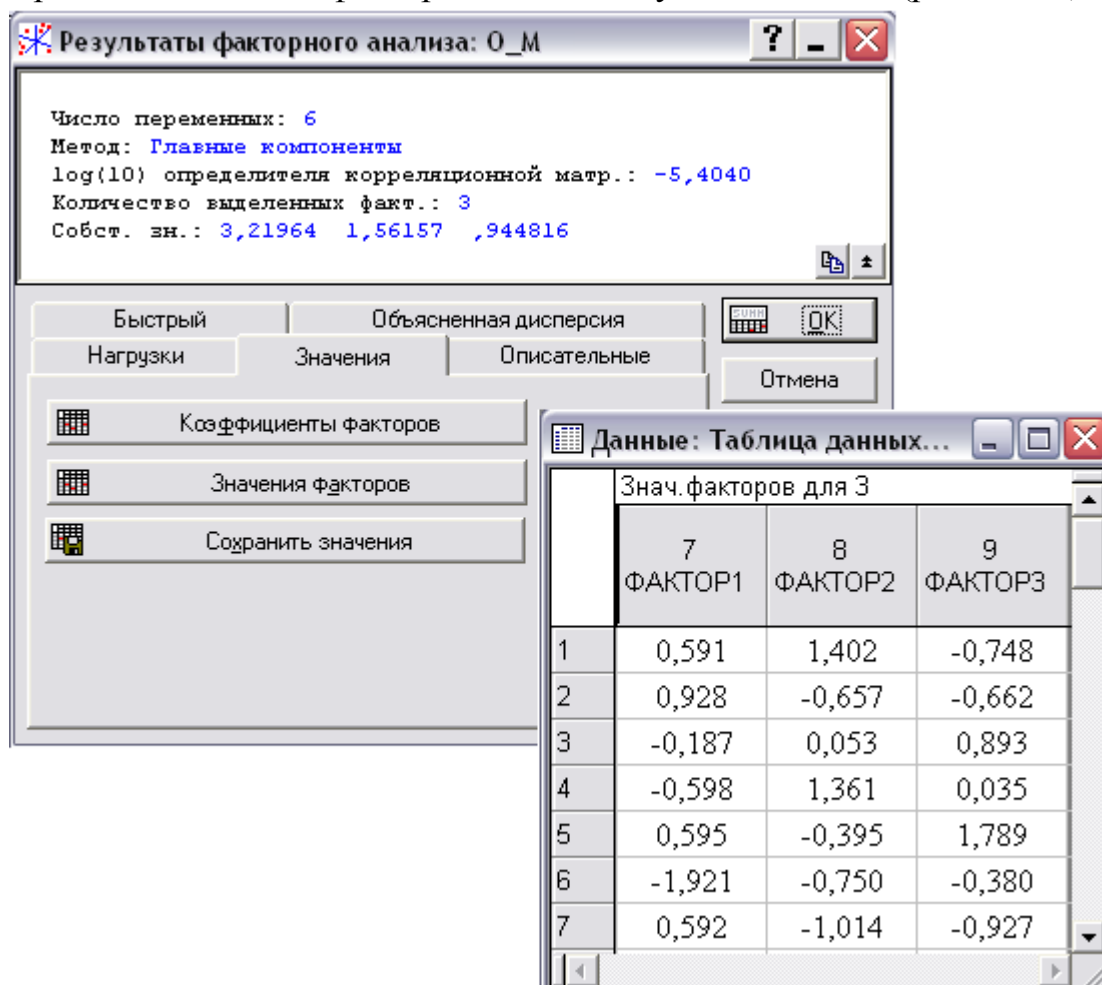



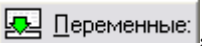



Рис. 3.200. Значения новых (факторных) переменных по каждому наблюдению для 3-х факторной модели выборки T*I

3.3.8. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов в T*I.

Шаг 1. Запустите в головном меню модуль **Анализ**. Выберите **Многомерный разведочный анализ** в стартовой панели модуля и далее щелкните мышкой на разновидности **Кластерный анализ**. В появившемся окне выде-

лите разновидность  и щелкните  (рис. 3.201).

Шаг 2. В открывшемся окне  нажмите кнопку , выберите все факторные переменные (рис. 3.201) и нажмите .

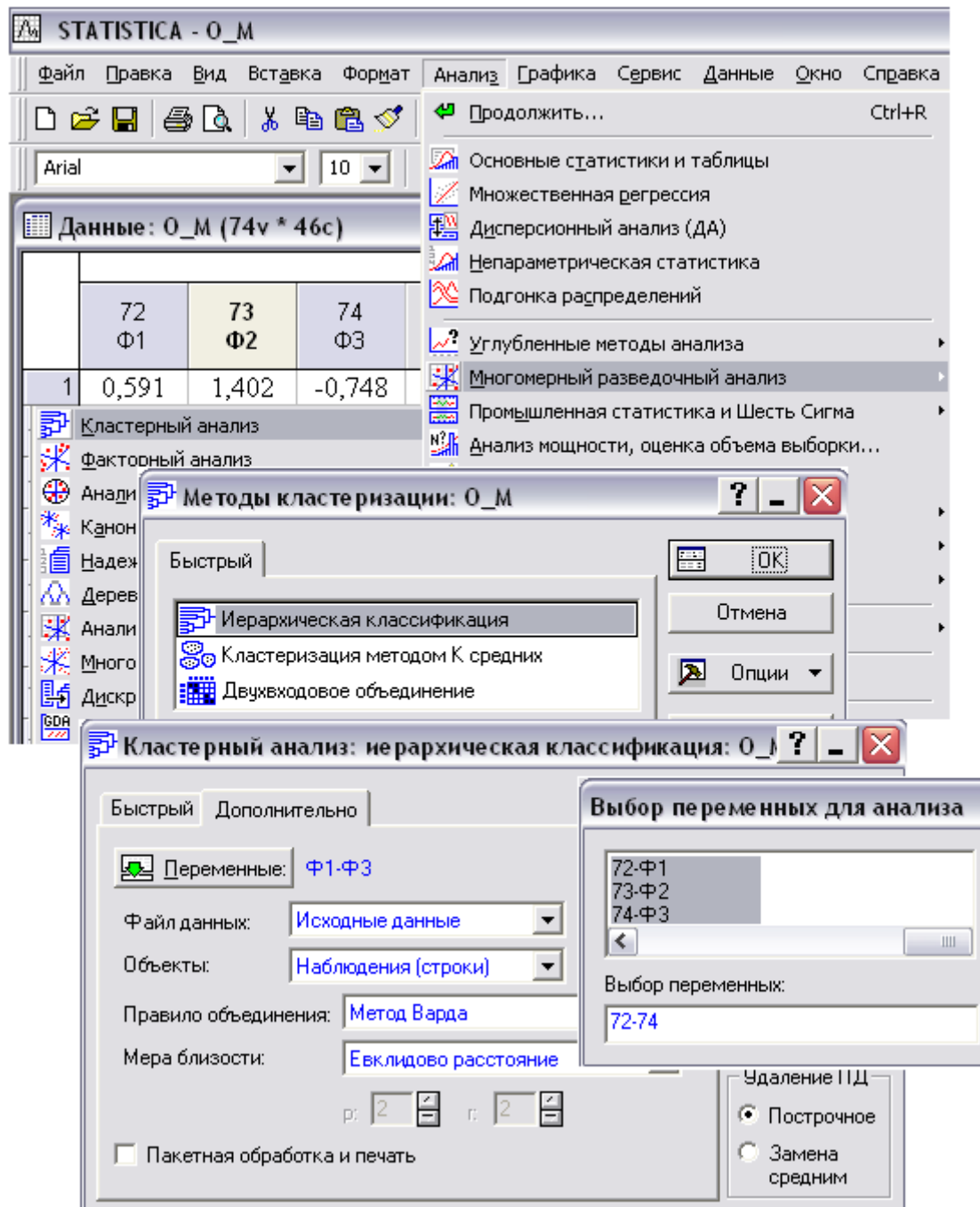



Рис.3.201. Окно  для факторных наблюдений Т*І:



Шаг 3. Далее в окне **Объекты:** выделите **Наблюдения (строки)**, в окне **Мера близости:** выделите **Евклидово расстояние**, а в окне **Правило объединения:** — **Метод Варда** (рис. 3.201) и нажмите **OK**. После этого щелкните **Вертикальная дендрограмма** (рис. 3.202). Следующая дендрограмма наблюдений появится на экране.

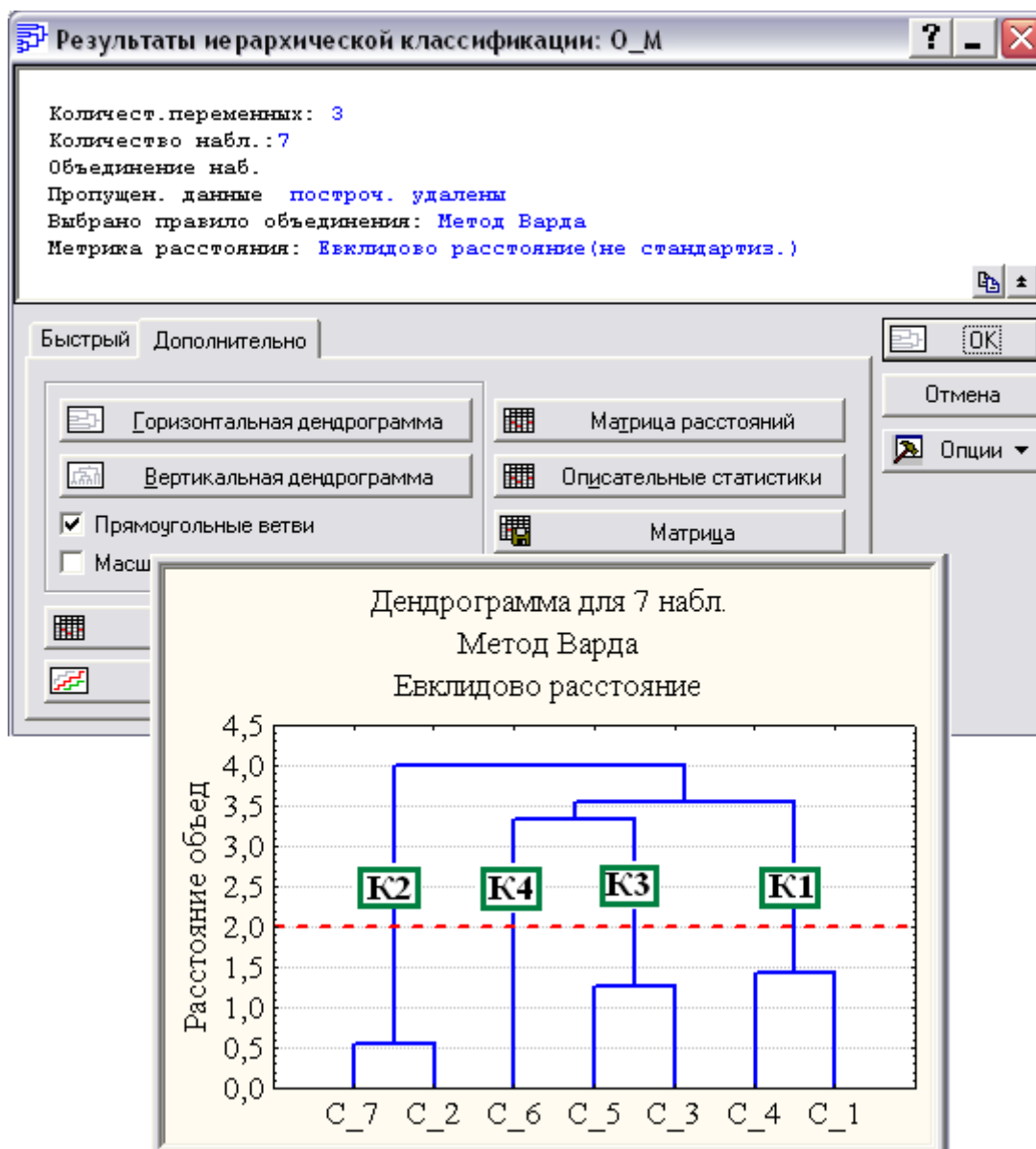


Рис. 3.202. Окно **Результаты иерархической классификации:** наблюдений T*I

Таким образом, в построенном 3-х мерном факторном пространстве $\{\Phi_1, \Phi_2, \Phi_3\}$ проведена кластеризация наблюдений. При этом в

качестве меры близости выбрано расстояние Евклида, а в качестве правила объединения двух кластеров использован дисперсионный по сути метод Уорда (Варда). Методом древовидной кластеризации построено иерархическое дерево (рис. 3.202), стабильно делящее 7 наблюдений на 4 кластера (K1: C1+C4; K2: C2+C7; K3: C3+C5 K4: C6) по отношению к вариации как к **Мера близости:** (расстояние Евклида, манхеттенское, Чебышева), так и к **Правило объединения:** (метод Варда, полной связи, взвешенное попарное среднее).

Наряду с методом древовидной кластеризации применяется также метод *K*-средних, проводящий классификацию объектов (наблюдений) по заданному количеству кластеров. Алгоритм метода *K*-средних, перемещая объекты в разные кластеры с целью минимизации изменчивости внутри кластеров и максимизации изменчивости между кластерами, оценивает качество кластеризации наблюдений по каждому фактору посредством параметрического дисперсионного анализа.

Шаг 4. В окне **Методы кластеризации:** (рис. 3.201) выберите метод **Кластеризация методом K-средних** и щелкните **OK**. В открывшемся окне нажмите кнопку **Переменные:**, выберите все факторные переменные аналогично рис. 3.201 и нажмите **OK**. В окне **Объекты:** выделите **Наблюдения (строки)**, в окне **Число кластеров:** выделите 4 в соответствии с рис. 3.202, а в окне **Начальные центры кластеров** – **Сортировать расстояния** (рис.3.203) и нажмите **OK** для просмотра результатов кластеризации наблюдений методом *K*-средних (рис. 3.203). Метод *K*-средних является итеративной процедурой. Алгоритм, реализованный в **Кластерный анализ**, весьма эффективен и обычно нет необходимости менять установку по умолчанию (10 итераций) в поле ввода **Число итераций:**.

Заметим, что опции в окне **Начальные центры кластеров** позволяют выбрать способ вычисления начального расположения центров кластеризации, от которого до некоторой степени зависят результаты метода *K*-средних, если имеется много небольших кластеров, которые ясно различимы. Опция **«Максимизация расстояний между кластерами»** выбирает перебором наблюдения в качестве начального расположения кластеров так, чтобы максимизировать начальные кластерные расстояния. Процедура может дать кластеры с единственным наблюдением, если они явно выделяются среди данных. Если выбрана опция **«Сортировать расстояния и взять наблюдения на постоянных интервалах»**,

то сначала сортируются расстояния между всеми объектами, а затем в качестве начальных кластерных центров выбираются объекты на постоянных интервалах. Опция «**Выбрать первых N (количество кластеров) наблюдений**» выбирает первые N (количество кластеров) наблюдений в качестве начальных кластерных центров, тем самым давая пользователю полный контроль над выбором начальной конфигурации и позволяя перемещать в начало файла наблюдения, которые вы хотите выбрать в качестве начальных кластерных центров.

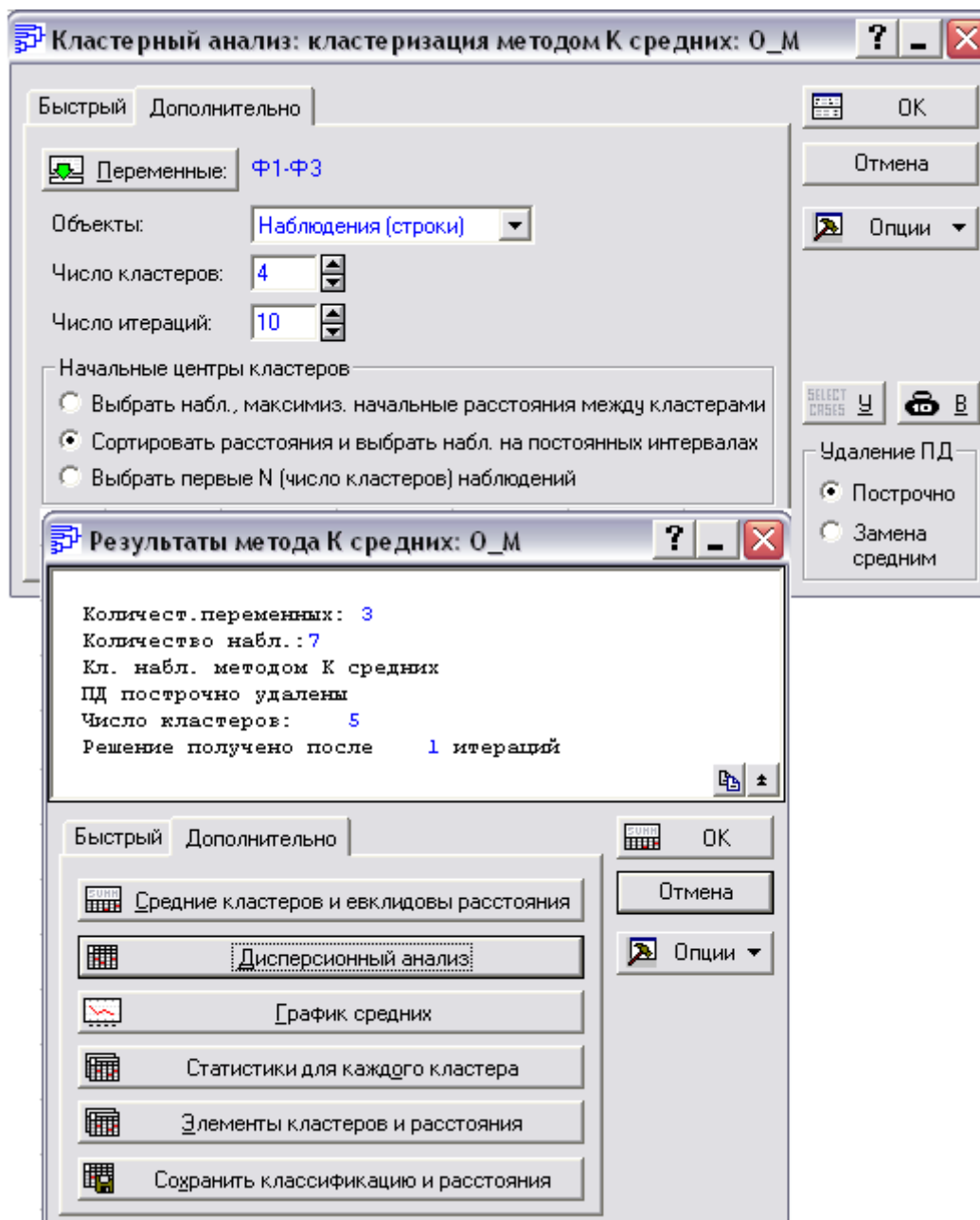



Рис. 3.203. Окно  Кластеризация методом К средних наблюдений Т*І


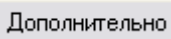
Рассмотрим результаты кластеризации наблюдений методом K -средних (рис. 3.203)

Шаг 5. Нажмите прежде всего кнопку  Дисперсионный анализ для построения стандартной таблицы результатов дисперсионного анализа кластеризации наблюдений по каждому фактору (рис. 3.204).

перемен.	Дисперсионный анализ (O_M)					
	Между SS	сс	Внутри SS	сс	F	значим. p
Ф1	4,931	3	1,069	3	4,612	0,12062
Ф2	5,835	3	0,165	3	35,366	0,00768
Ф3	5,257	3	0,743	3	7,079	0,07112

Рис. 3. 204. Результаты дисперсионного анализа кластеризации наблюдений $T \times I$ по факторам


Признаком хорошей классификации является малость изменчивости внутри кластера в сравнении с изменчивостью вне кластера, что оценивается с помощью стандартного межгруппового дисперсионного анализа для каждого фактора. Значения F -статистики, полученные для каждого фактора, является индикатором того, насколько хорошо соответствующий фактор дискриминирует кластеры. Значения p -уровней определяют статистически значимым (на уровне значимости $0,005 < p \approx 0,00768 < 0,050$) различия между кластерами для фактор Ф2, слабо значимым (на уровне значимости $0,05 < p \approx 0,071 < 0,10$) различия между кластерами для фактора Ф3 и почти незначимым (на уровне значимости $0,10 < p \approx 0,12$) различия между кластерами для фактор Ф1.

Шаг 6. Нажмите кнопку  Средние кластеров и евклидовы расстояния в режиме  Дополнительно для вывода на экран таблицы результатов со средними для каждого кластера по каждому фактору (рис. 3.205).

перемен.	Средн.класт. (O_M)			
	Кластер No. 1	Кластер No. 2	Кластер No. 3	Кластер No. 4
Ф1	-1,921	0,204	-0,004	0,760
Ф2	-0,750	-0,171	1,382	-0,836
Ф3	-0,380	1,341	-0,357	-0,794

Рис. 3.205. Кластерные средние наблюдений $T \times I$ в факторном пространстве

Здесь же доступна просмотру таблица результатов с евклидовыми расстояниями между "центрами кластеров". В данном случае матрица показывает евклидовы расстояния между кластерами, вычисленные из средних значений соответствующих кластеров по факторам, используемым для классификации.

Шаг 6'. Нажмите кнопку  **График средних** для вывода на экран графиков факторных средних для каждого кластера (рис. 3.206).

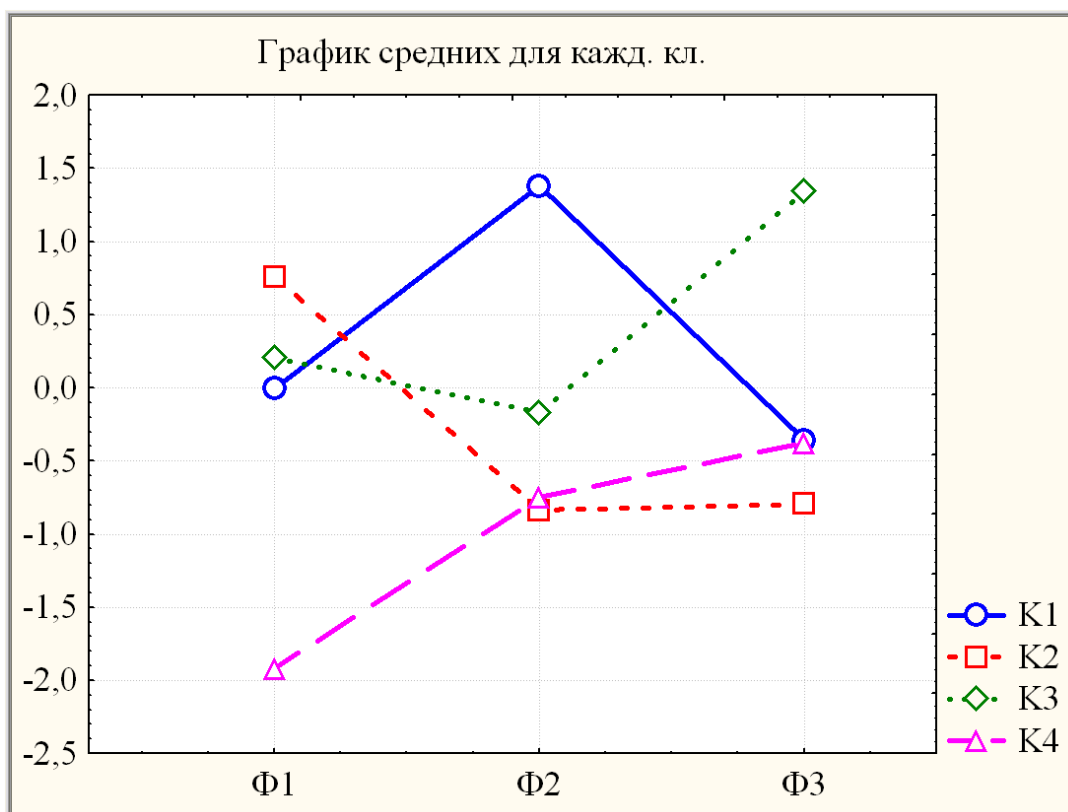

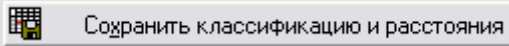


Рис. 3. 206. Линейные графики факторных средних наблюдений $T \times I$ для каждого кластера

Заметим, что с помощью  **Элементы кластеров и расстояния** (рис. 3. 203) можно вывести на экран списки элементов (наблюдений) каждого кластера, где для каждого наблюдения вычислено также его расстояние до центра кластера. Полученные результаты кластеризации наблюдений методом K -средних являются устойчивыми по отношению к вариации опций в окне **Начальные центры кластеров**: («Максимизация расстояний между кластерами», «Сортировать расстояния и взять наблюдения на постоянных интервалах», «Выбрать первых N (количество кластеров) наблюдений») и в полной мере соответствуют иерархическому дереву (рис. 3.202), построенному методом древовидной кластеризации и

стабильно делящему 7 наблюдений на 4 кластера (K1: C1+C4; K2: C2+C7; K3: C3+C5 K4: C6).

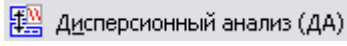
Шаг 7. Нажмите  для вывода на экран данных о классификации и расстояниях наблюдений до центра соответствующего кластера, в дополнение к факторным переменным, выбранным из исходного файла данных (рис. 3.207).

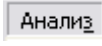
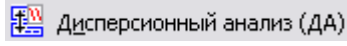
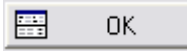
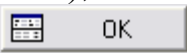


	O_M					
	1 Ф1	2 Ф2	3 Ф3	4 НАБЛ. №	5 КЛАСТЕР	6 РАССТ.
C_1	0,591	1,402	-0,748	1	1	0,41
C_2	0,928	-0,657	-0,662	2	2	0,16
C_3	-0,187	0,053	0,893	3	3	0,37
C_4	-0,598	1,361	0,035	4	1	0,41
C_5	0,595	-0,395	1,789	5	3	0,37
C_6	-1,921	-0,750	-0,380	6	4	0,00
C_7	0,592	-1,014	-0,927	7	2	0,16

Рис. 3.207. Результаты факторной классификации наблюдений T*I

В связи с дальнейшей проверкой построенных кластеров на однородность по каждому фактору оформим стандартную процедуру дисперсионного анализа результатов факторной классификации наблюдений.

Продемонстрируем возможности модуля  для расширенного анализа полученного результата. Так как значения одномерного F -критерия для двух факторов – значимы (рис. 3.204), то далее можно использовать метод множественных сравнений для выявления тех групп кластеров для каждого фактора, которые вызвали данный эффект и максимально отличаются от других, а также можно выделить однородные (различающиеся незначимо, то есть $p > 0,10$) иногда пересекающиеся группы кластеров.

Шаг 8. Запустите  и щелкните  (см. рис. 3.208). В появившемся окне выберите тип анализа (однофакторный дисперсионный анализ) и задайте метод (диалог быстрых спецификаций), как показано на рис.3.208, и нажмите . В появившемся окне выберите в качестве зависимых все факторные переменные (позиции 72-74), а в качестве категориальной КЛАСТЕР (позиция 75), закажите все коды (1-4), как показано на рис. 3.208, и нажмите .

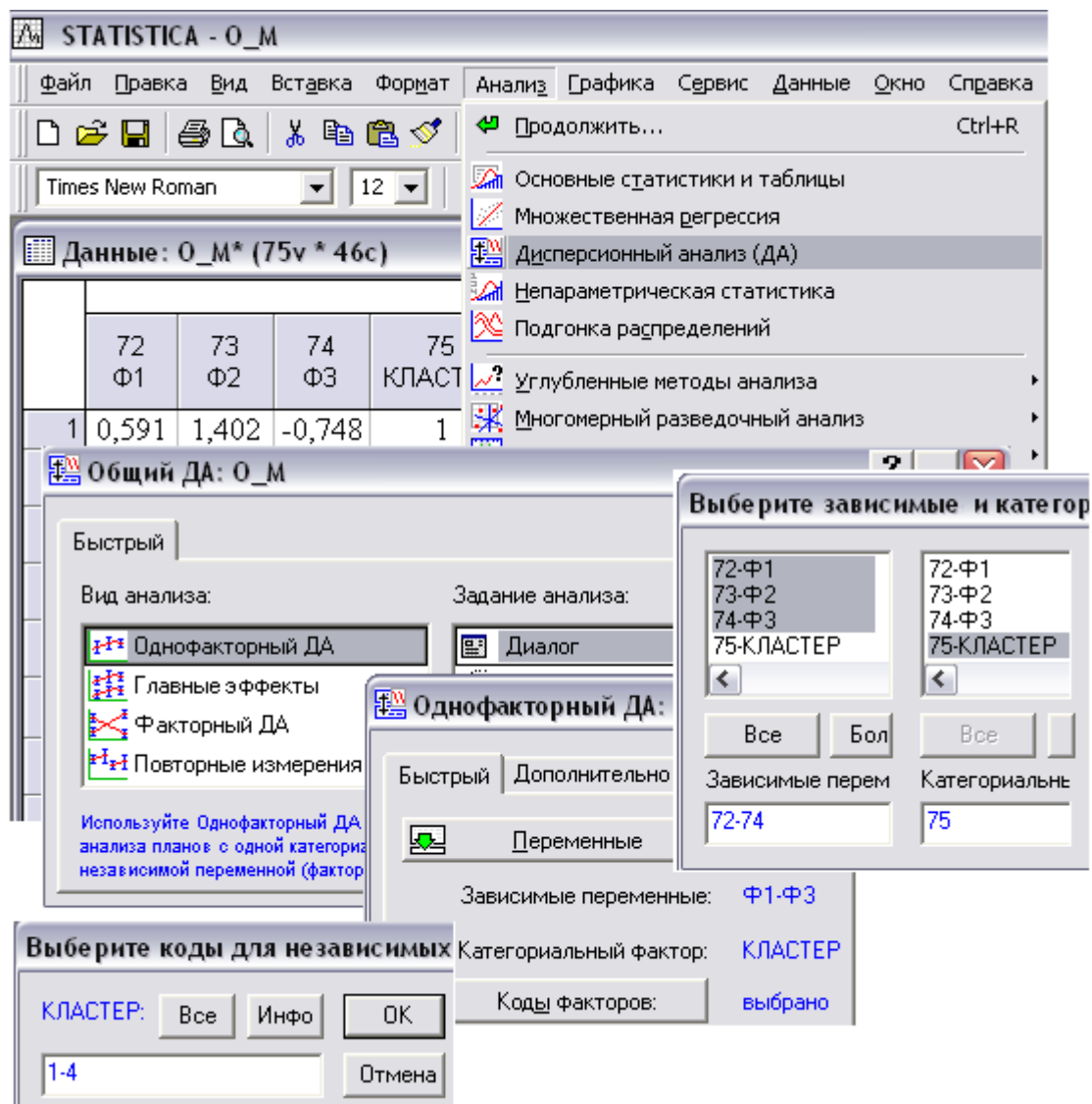


Рис. 3.208. Выбор переменных для дисперсионного анализа кластеров наблюдений Т*І

Шаг 9. После этого в появившемся окне **Результаты анализа** (рис.3.66) в режиме **Итоги** нажмите **Одномерн. результаты** для просмотра результатов дисперсионного анализа (рис. 3.204).

Шаг 10. В окне **Результаты анализа** (рис. 3.66) нажмите кнопку **Больше** и в открывшемся окне выберите режим **Апостериорные**, в котором представлены различные апостериорные критерии (рис. 3.209). Выберите, например, критерий **Фишера НЗР** с опцией отображения **Значимые различия** (рис. 3. 209).

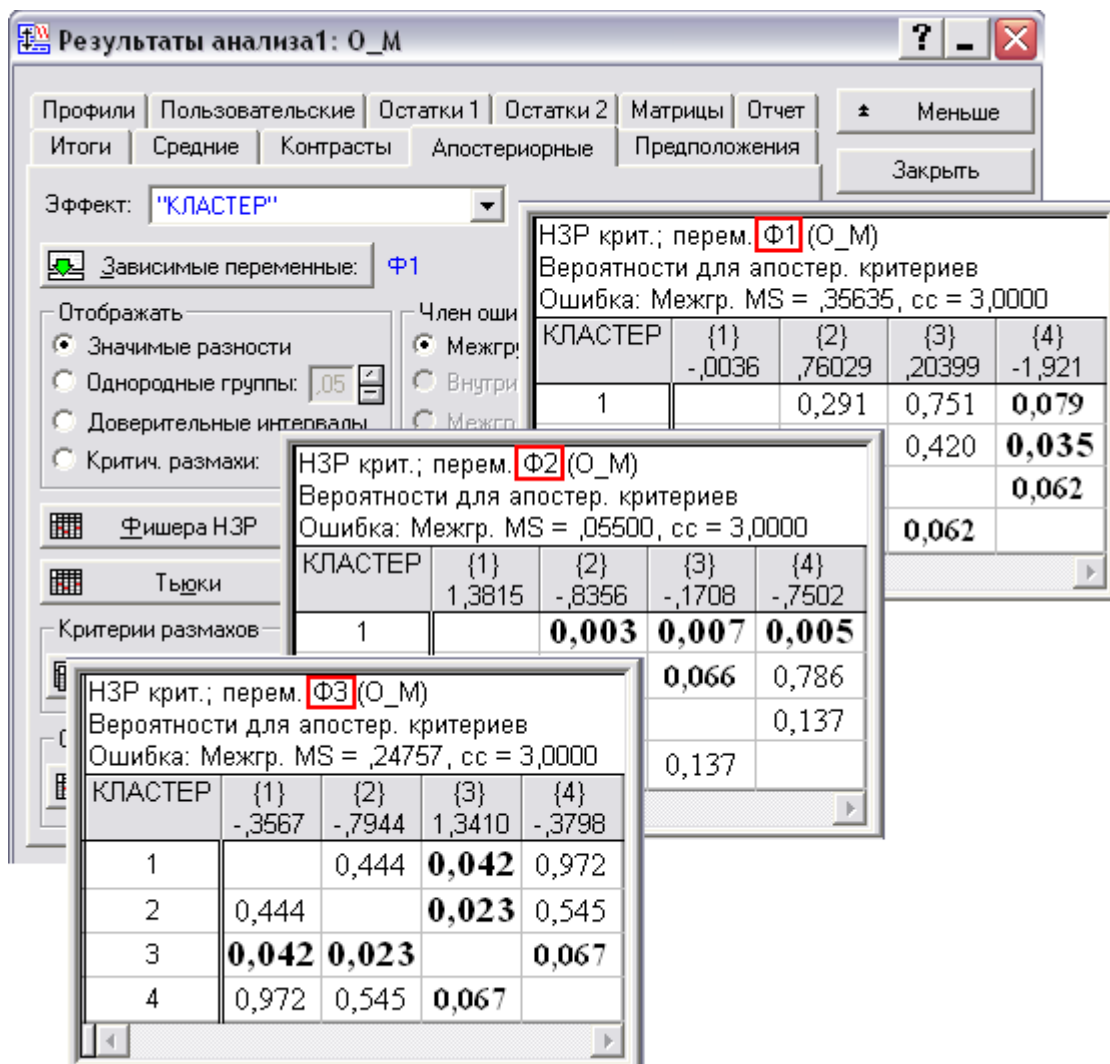


Рис. 3.209. Режим **Апостериорные** для выбора апостериорного критерия множественных сравнений средних **Фишера НЗР**

Согласно рис. 3.209, для каждого фактора по апостериорному критерию **Фишера НЗР** выделены моно-кластер и однородная группы из трех оставшихся кластеров. Заметим, что критерий НЗР дает самые значимые результаты по сравнению с другими апостериорными критериями.

Таким образом, по апостериорному критерию НЗР:

Ф1: {K2, K3, K1} и {K4} так, что K4 ($m_{K4} \approx -1,92$) отличается от K2 ($m_{K2} \approx 0,76$) статистически значимо на уровне значимости $p \approx 0,035$, а от K3 ($m_{K3} \approx 0,20$) и K1 ($m_{K1} \approx -0,004$) – слабо значимо ($0,10 > p > 0,05$).

Ф2: {K1} и {K3, K4, K2}, так, что K1 ($m_{K1} \approx 1,38$) отличается от K3 ($m_{K3} \approx -0,17$) статистически значимо на уровне значимости $p \approx 0,007$,

а от К2 ($m_{K2} \approx -0,84$) даже сильно значимо на уровне значимости $p \approx 0,003$.

Ф3: {К3} и {К1, К4, К2} так, что К3 ($m_{K3} \approx 1,34$) отличается от К1 ($m_{K1} \approx -0,36$) и от К2 ($m_{K2} \approx -0,79$) статистически значимо ($0,05 > p > 0,005$), а от К4 ($m_{K4} \approx -0,38$) – слабо значимо и на уровне значимости $p \approx 0,067$.

Результаты кластерного анализа наблюдений по совокупности факторов (рис. 3.126-127) с учетом результатов сравнений кластерных средних для каждого фактора позволяют провести классификацию наблюдений в порядковой шкале стандартизированных измерений, полагая в качестве уровня «Средний» – стандартизированный интервал (-0,5; +0,5) наблюдений, «Выше среднего» – (+0,5; +1,5) наблюдений, «Ниже среднего» – (-1,5; -0,5). Аномально низкие значения ($< -1,5$) определяют уровень «Аутсайдер».

Таблица 3.32

Классификация наблюдений Т*І по совокупности факторов в порядковой шкале стандартизированных измерений.

Кластер	Уровень кластера на фоне среднего по фактору			Состав кластера
	Ф1 (Na – Cr)	Ф2 (Ca)	Ф3 (Ba – Sr – Fe)	
К1	Средний	Выше среднего	Средний	C_1+4
К2	Выше среднего	Ниже среднего	Ниже среднего	C_2+7
К3	Средний	Средний	Выше среднего	C_3+5
К4	Аутсайдер	Ниже среднего	Средний	C_6

Для достоверности полученных результатов параметрического дисперсионного анализа необходимо проверить предположения о нормальном распределении сравниваемых кластеров и об однородности дисперсий в них. Но в данном случае достаточно замечания о малочисленности состава кластеров, чтобы обратиться к непараметрическому однофакторному дисперсионному анализу на основе критерия Краскела-Уоллиса и медианного теста.

Шаг 11. Запустите модуль **Анализ**, щелкните в нем по кнопке **Непараметрическая статистика** (см. рис. 3.210), затем выберите тип сравнения **Сравнение нескольких независимых групп**, закажите переменные стандартным образом (см. рис. 3. 210) и щелкните по **ДА Краскела-Уоллиса и медианный тест** и **Сравнения средних рангов для всех групп** (рис. 3. 210).

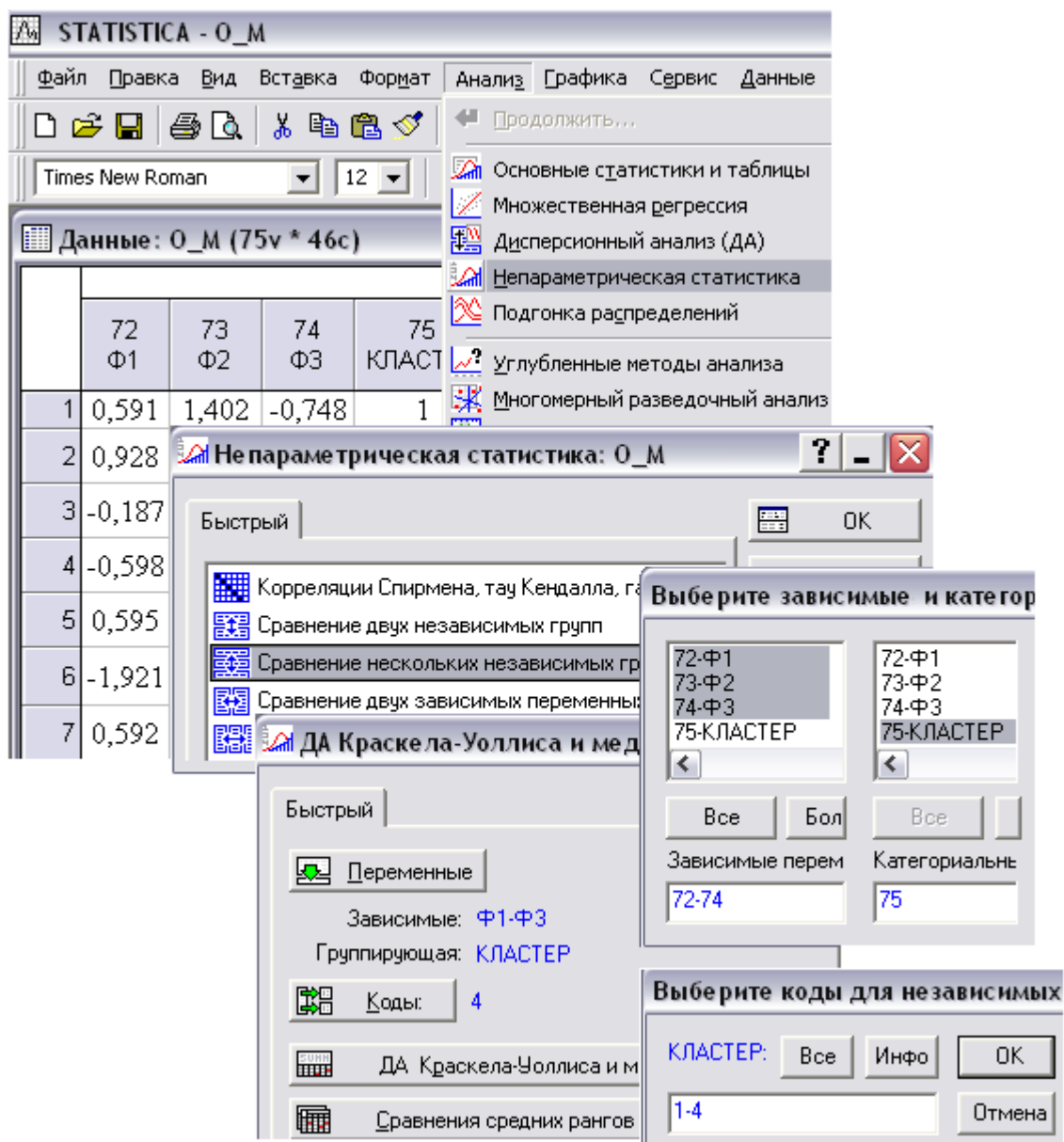


Рис. 3.210. **Непараметрическая статистика** в головном модуле **Анализ** для базы факторных данных T*I

В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа (рис. 3.211).

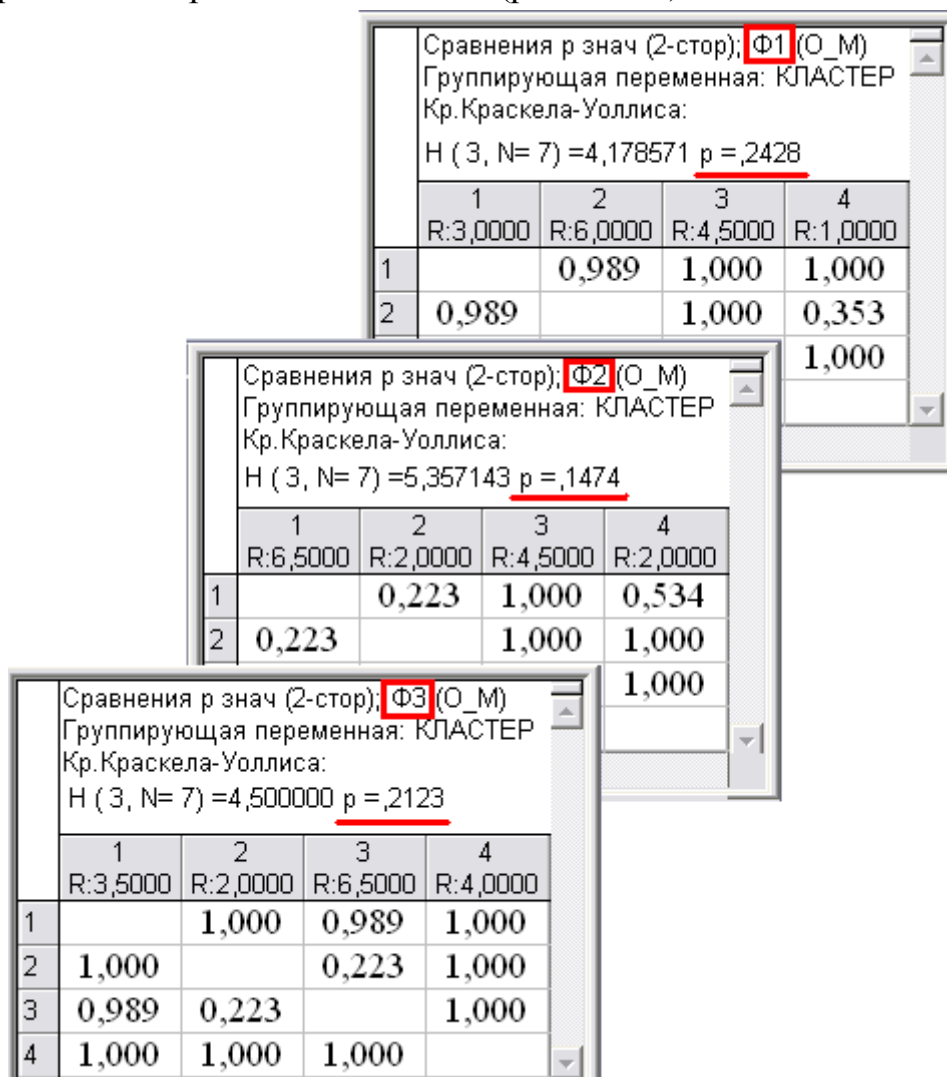


Рис. 3.211. Результаты рангового дисперсионного анализа кластеризации наблюдений Т*І на основе критерия Краскела-Уоллиса

По совокупности критерия Краскела-Уоллиса (рис. 3.211) и медианного теста 4 кластера К1-К4 различается незначимо по каждому фактору. Таким образом, непараметрический дисперсионный анализ существенно «смягчает» результаты параметрического о значимости кластеризации наблюдений выборки Т*І. В силу расхождения результатов дисперсионных анализов в связи с малочисленностью состава кластеров более корректными признаются результаты непараметрического дисперсионного анализа. Если все же использовать параметрический подход, то полученную статистическую значимость различий кластерных средних (рис. 3. 209) следует «смягчить» до слабой значимости.

3.3.9. Факторный анализ содержаний микроэлементов в накипи территории Ч, полученных методом ISP (выборка Ч*І).

Рассмотрение проведем аналогично п. 3.3.7.


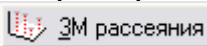
Шаг 1. Запустите , нажмите , где далее выберите аналогично рис. 3.101.

Шаг 2. Выберите (рис. 3.212) отдельной базы данных Ч*І в открывшемся окне аналогично рис. 3.187.

	71 Na ₂ С ₄ І	72 Са ₂ С ₄ І	73 Fe ₂ С ₄ І	74 Sr ₂ С ₄ І	75 Ba ₂ С ₄ І	76 Сг ₂ С ₄ І
1	1,830	25,5	0,40	0,217	0,010	27,0
2	0,890	36,9	0,44	0,361	0,014	16,0
3	2,480	21,7	0,27	0,194	0,011	38,0
4	2,330	3,6	0,02	0,029	0,001	41,0
5	1,050	34,5	0,70	0,281	0,012	19,0
6	2,690	37,5	0,11	0,203	0,002	46,0
7	2,400	34,1	0,05	0,173	0,009	35,0
8	0,630	37,3	0,02	0,196	0,000	11,0
9	1,630	36,8	0,03	0,096	0,006	25,0
10	0,620	40,5	0,10	0,202	0,005	3,0
11	1,510	38,9	0,10	0,245	0,009	23,0
12	0,790	39,0	0,11	0,266	0,013	16,0
13	0,370	37,0	0,14	0,244	0,010	8,0
14	1,600	41,4	0,09	0,169	0,005	18,0
15	2,260	41,1	0,43	0,262	0,010	30,0

Рис. 3. 212. Выбор переменных выборка Ч*І для факторного анализа

Шаг 3. Нажмите кнопку в режиме аналогично рис. 3.188. Можно посмотреть аналогично рис. 3.189 в режиме уже знакомую корреляционную матрицу (см. табл.3.29) и прочую описательную статистику выборки (и), построить различные графики,

например,  2М рассеяния Na – Cr (рис. 3.213) аналогично рис. 3.190 или  3М рассеяния Ba – Sr – Fe (рис. 3.214) аналогично рис. 3.191.

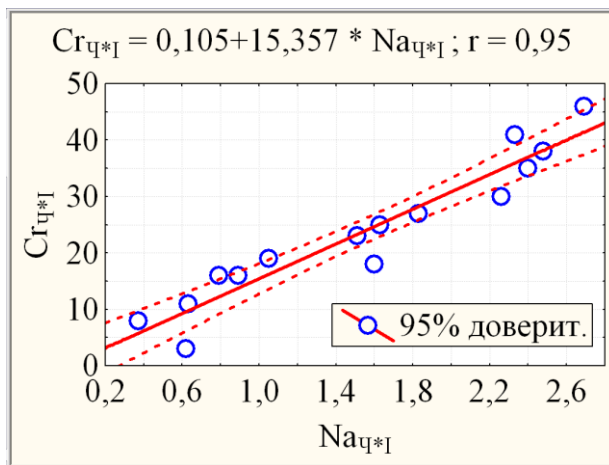


Рис.3.213. 2М Диаграмма рассеяния $Na_{Ч*I} - Cr_{Ч*I}$

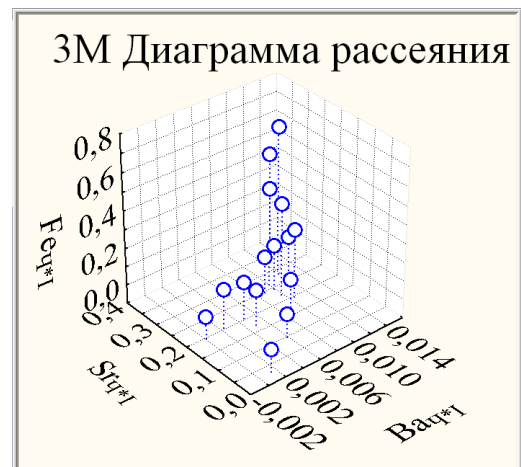
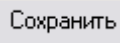



Рис.3.214. 3М Диаграмма рассеяния $Ba_{Ч*I} - Sr_{Ч*I} - Fe_{Ч*I}$

Шаг 4. Описательную статистику можно посмотреть также в режиме  Сохранить (см. рис. 3.107). Нажмите  Матрицу для вывода на экран корреляционной матрицы, расширенной строками средних и стандартных отклонений (рис. 3.215) аналогично рис. 3.192.

Данные: Таблица данных72* (6v * 10с)

	O_M					
	1	2	3	4	5	6
	$Na_{Ч*I}$	$Ca_{Ч*I}$	$Fe_{Ч*I}$	$Sr_{Ч*I}$	$Ba_{Ч*I}$	$Cr_{Ч*I}$
$Na_{Ч*I}$	1,000	-0,431	-0,051	-0,432	-0,168	0,954
$Ca_{Ч*I}$	-0,431	1,000	0,053	0,577	0,208	-0,536
$Fe_{Ч*I}$	-0,051	0,053	1,000	0,630	0,642	-0,041
$Sr_{Ч*I}$	-0,432	0,577	0,630	1,000	0,708	-0,412
$Ba_{Ч*I}$	-0,168	0,208	0,642	0,708	1,000	-0,156
$Cr_{Ч*I}$	0,954	-0,536	-0,041	-0,412	-0,156	1,000
Средние	1,539	33,715	0,201	0,209	0,008	23,733
Стд. откл	0,777	9,985	0,202	0,078	0,004	12,510
Кол-во N	15,000					

Рис. 3. 215. Наблюдаемая корреляционная матрица, расширенная строками средних и стандартных отклонений выборки Ч*I

Шаг 5. Для проведения факторного анализа выборки вернитесь в окно **Задайте метод выделения факторов** аналогично рис. 3.188 с помощью кнопки **Отмена** и в режиме **Дополнительно** в поле справа укажите Максимальное число факторов, которые будут выделены системой, например, в начале исследования 6, и Минимальное собственное значение, например 0,00 аналогично рис. 3.193. В окне **Метод выделения** выберите в качестве метода анализа **Главные компоненты** и щелкните **OK**. На экране появятся результаты анализа аналогично рис. 3.194. Например, для просмотра расширенной информации о собственных значениях в режиме **Объясненная дисперсия** щелкните по кнопке **Собственные значения**. Дисперсия, объясненная последовательностью собственных значений, представит в форме рис. 3.216.

Значен.	Собственные значения (O_M)			
	Соб. зн.	% общей дисперс.	Кумулятивн. собст. знач.	Кумулятивн. %
1	3,060	51,006	3,060	51,006
2	1,730	28,829	4,790	79,835
3	0,688	11,465	5,478	91,300
4	0,341	5,678	5,819	96,978
5	0,152	2,538	5,971	99,516
6	0,029	0,484	6,000	100,000

Рис. 3.216. Результаты факторного анализа для выборки Ч*І

В первом столбце (рис. 3.216) приведены собственные значения, во втором - % от общей дисперсии для каждого собственного значения, в третьем – накопленные собственные значения и в четвертом – их % от общей дисперсии. Таким образом, первое собственное значение объясняет $\approx 51,01\%$ от общей дисперсии, совокупность первых двух собственных значений – уже $\approx 79,84\%$, а совокупность первых трех собственных значений объясняет $\approx 91,30\%$ от общей дисперсии.

Для проведения графического «теста каменистой осыпи» выполните следующий шаг:

Шаг 6. В режиме **Объясненная дисперсия** (рис. 3.194) щелкните по кнопке **График каменистой осыпи** для просмотра графической информации о всех собственных значениях (рис. 3.217) аналогично рис. 3.195.

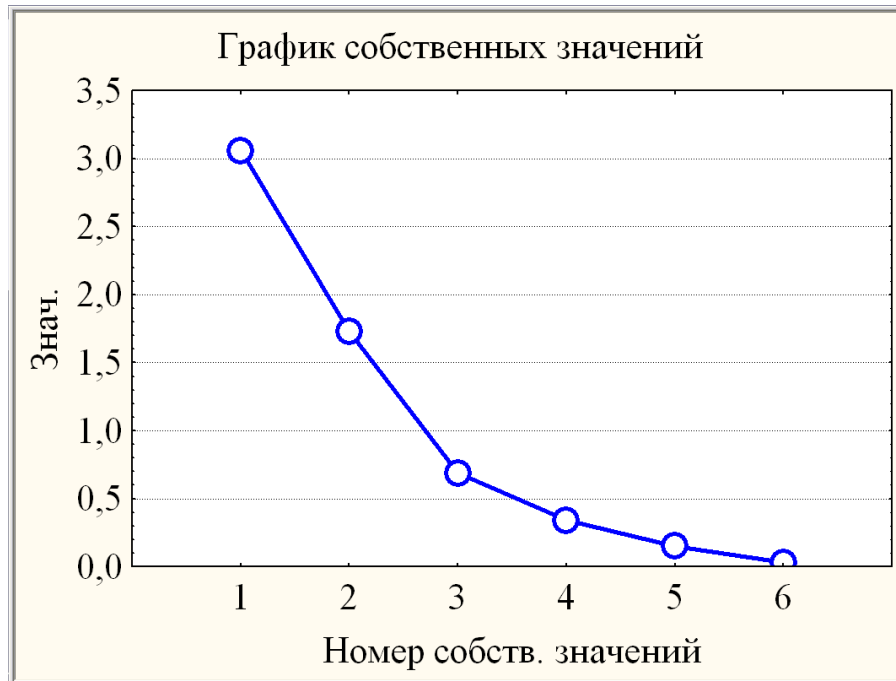


Рис. 3.217. График собственных значений для выборки Ч*Г

Рис. 3. 217 демонстрирует графический метод («критерий каменистой осыпи») оценки значимости собственных значений: справа от критической точки, где убывание собственных значений слева направо максимально замедляется, находится только «факториальная осыпь», которой пренебрегают. В соответствии с этим критерием можно оставить 3 фактора.

Самую важную информацию, на которой базируется интерпретация факторов, содержат факторные нагрузки, которые могут интерпретироваться как корреляции между факторами и переменными. Посмотрим на вращаемые факторные нагрузки в 3-х факторной модели:

Шаг 7. Щелкните кнопкой , чтобы еще раз возвратиться в окно аналогично рис. 3.193 в режиме , в котором в поле справа укажите Максимальное число факторов, равное 3 и щелкните . На экране появятся результаты анализа 3-х факторной модели (рис. 3.218) в режиме , где выберите поворот варимаксимальный и щелкните по кнопке для просмотра факторных нагрузок в выбранной повернутой 3-х факторной модели (рис. 3.218) аналогично рис. 3.196.

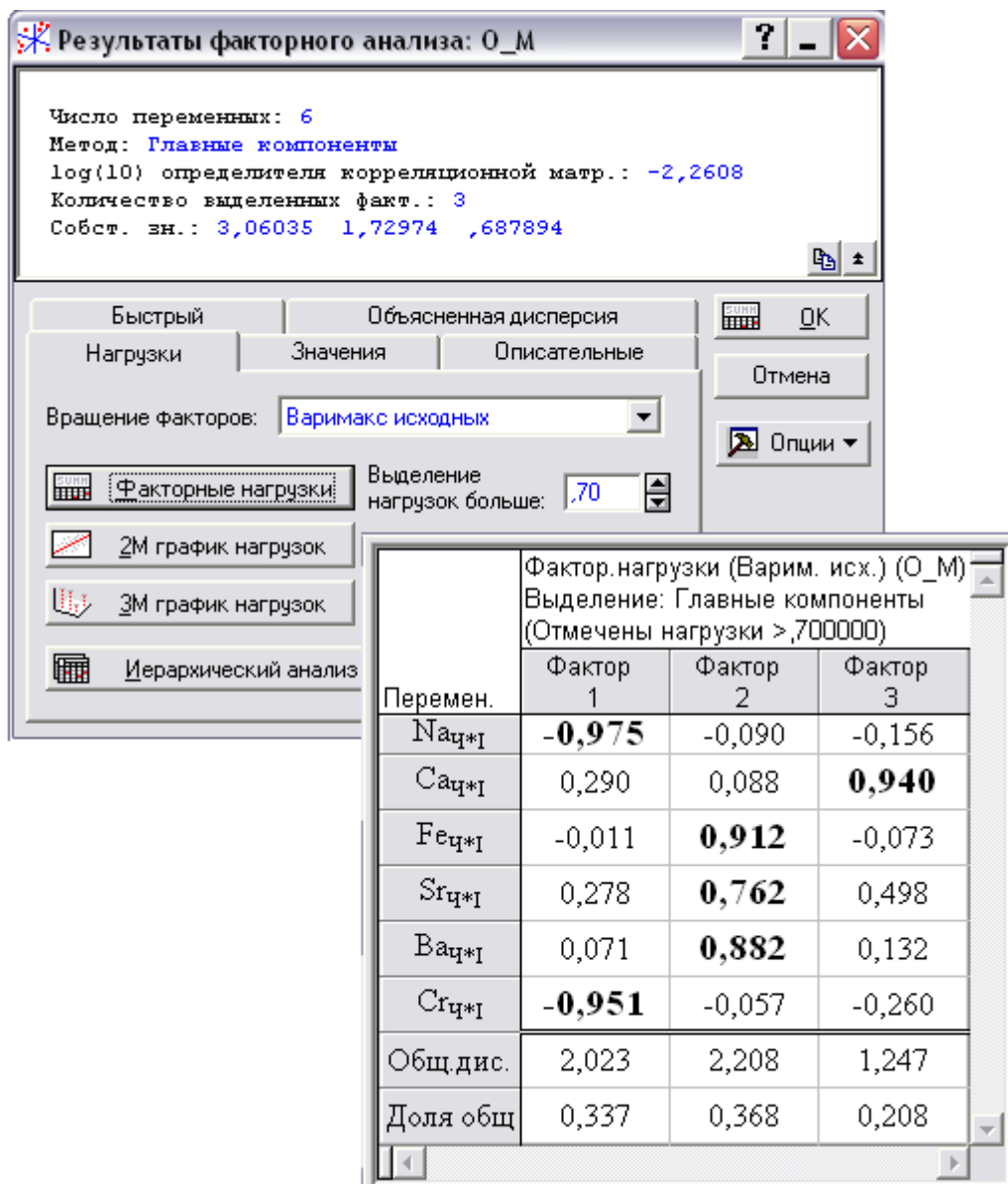


Рис. 3.218. Режим в окне для выборки Ч*І


Для предания факторам простой интерпретации пространство факторов подвергнуто стандартному вычислительному методу варимакс - вращения так, что факторы отмечены высокими нагрузками для одних переменных и низкими нагрузками для других.

Согласно рис. 3. 218, расположенные в порядке убывания высокие факторные нагрузки микроэлементов распределились по факторам следующим образом:

Фактор Ф2 – наиболее весомый, составной, характеризуется прежде всего микроассоциацией Fe – Ba, а также микроэлементом Sr, связанных с Ф2 положительной корреляционной связью.

Фактор Ф1 – менее весомый, составной, характеризуется микроассоциацией Na – Cr, связанной с Ф1 отрицательной корреляционной связью.

Фактор Ф3 – еще менее весомый, характеризуется в основном микроэлементом Ca, связанным с Ф3 положительной корреляционной связью.

Шаг 8. Для геометрической интерпретации факторной модели щелкните по кнопке  ЭМ график нагрузок (рис. 3.219) аналогично рис. 3.197.

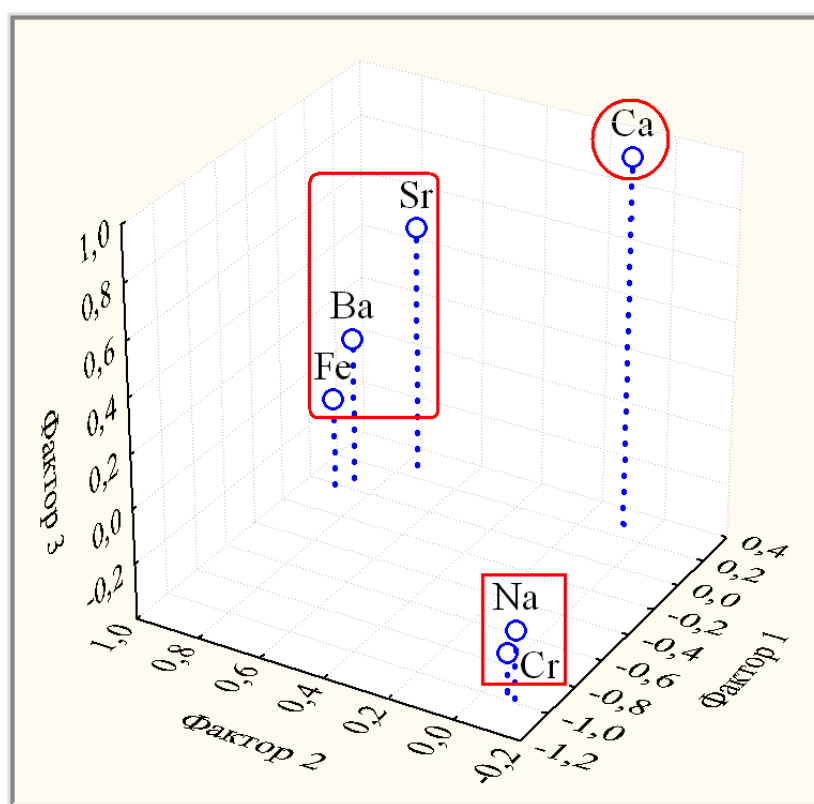

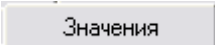
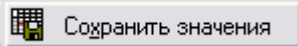
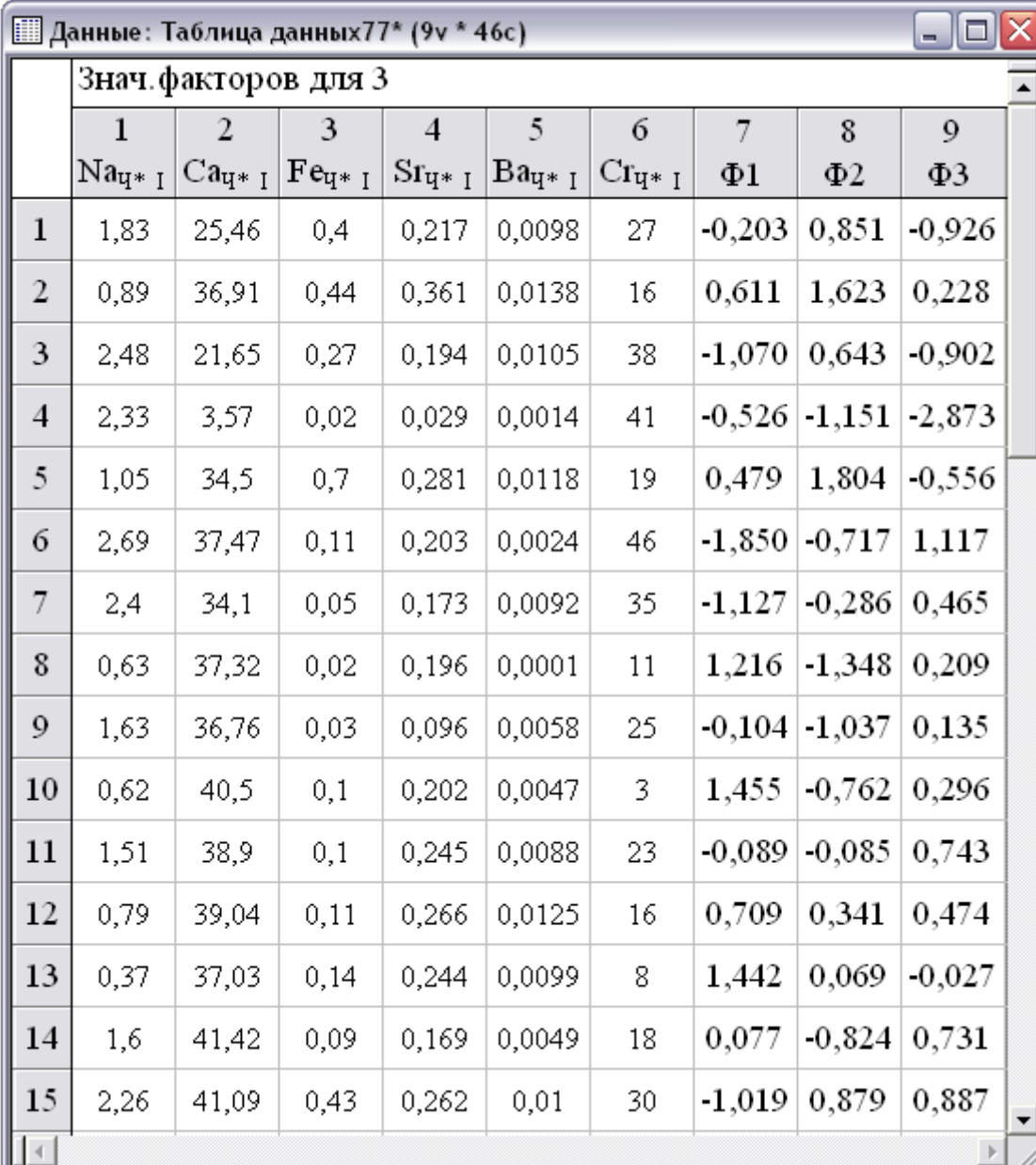


Рис. 3.219. Диаграмма рассеяния факторных нагрузок микроэлементов выборки Ч*І в координатах {Ф1, Ф2, Ф3}

На рис. 3.219 изображена 3-х мерная диаграмма рассеяния факторных нагрузок микроэлементов выборки Ч*І в координатах {Ф1, Ф2, Ф3}, достаточно наглядно разделяющая 6 микроэлементов на 3 микроассоциации.

Для проведения дальнейшего статистического анализа выборки Ч*І в рамках построенной 3-х факторной модели следует сохранить значения наблюдений в новой системе факторных координат.

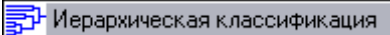

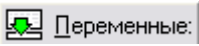



Шаг 9. В окне  **Результаты факторного анализа** войдите в режим  **Значения** аналогично рис. 3.200 и для сохранения значений факторов по каждому наблюдению щелкните по кнопке  **Сохранить значения**. На экране появится таблица значений факторов по каждому наблюдению наряду со старыми значениями наблюдений микроэлементов выборки Ч*І (рис. 3.220).



Знач. факторов для 3									
	1	2	3	4	5	6	7	8	9
	Na _ц * І	Ca _ц * І	Fe _ц * І	St _ц * І	Ba _ц * І	Cl _ц * І	Φ1	Φ2	Φ3
1	1,83	25,46	0,4	0,217	0,0098	27	-0,203	0,851	-0,926
2	0,89	36,91	0,44	0,361	0,0138	16	0,611	1,623	0,228
3	2,48	21,65	0,27	0,194	0,0105	38	-1,070	0,643	-0,902
4	2,33	3,57	0,02	0,029	0,0014	41	-0,526	-1,151	-2,873
5	1,05	34,5	0,7	0,281	0,0118	19	0,479	1,804	-0,556
6	2,69	37,47	0,11	0,203	0,0024	46	-1,850	-0,717	1,117
7	2,4	34,1	0,05	0,173	0,0092	35	-1,127	-0,286	0,465
8	0,63	37,32	0,02	0,196	0,0001	11	1,216	-1,348	0,209
9	1,63	36,76	0,03	0,096	0,0058	25	-0,104	-1,037	0,135
10	0,62	40,5	0,1	0,202	0,0047	3	1,455	-0,762	0,296
11	1,51	38,9	0,1	0,245	0,0088	23	-0,089	-0,085	0,743
12	0,79	39,04	0,11	0,266	0,0125	16	0,709	0,341	0,474
13	0,37	37,03	0,14	0,244	0,0099	8	1,442	0,069	-0,027
14	1,6	41,42	0,09	0,169	0,0049	18	0,077	-0,824	0,731
15	2,26	41,09	0,43	0,262	0,01	30	-1,019	0,879	0,887

Рис. 3.220. Значения новых (факторных) переменных по каждому наблюдению для 3-х факторной модели выборки Ч*І

3.3.10. Кластеризация наблюдений в рамках построенной факторной модели содержаний микроэлементов в Ч*І.

Шаг 1. Запустите разновидность  аналогично рис. 3.201. В открывшемся окне  нажмите кнопку , выберите все факторные переменные аналогично рис. 3.201 и нажмите . В окне **Объекты:** выделите , выберите разные сочетания **Мера близости:** и **Правило объединения:** (рис. 3.221). После этого щелкните  аналогично рис. 3.202. Следующие дендрограммы наблюдений появятся на экране.

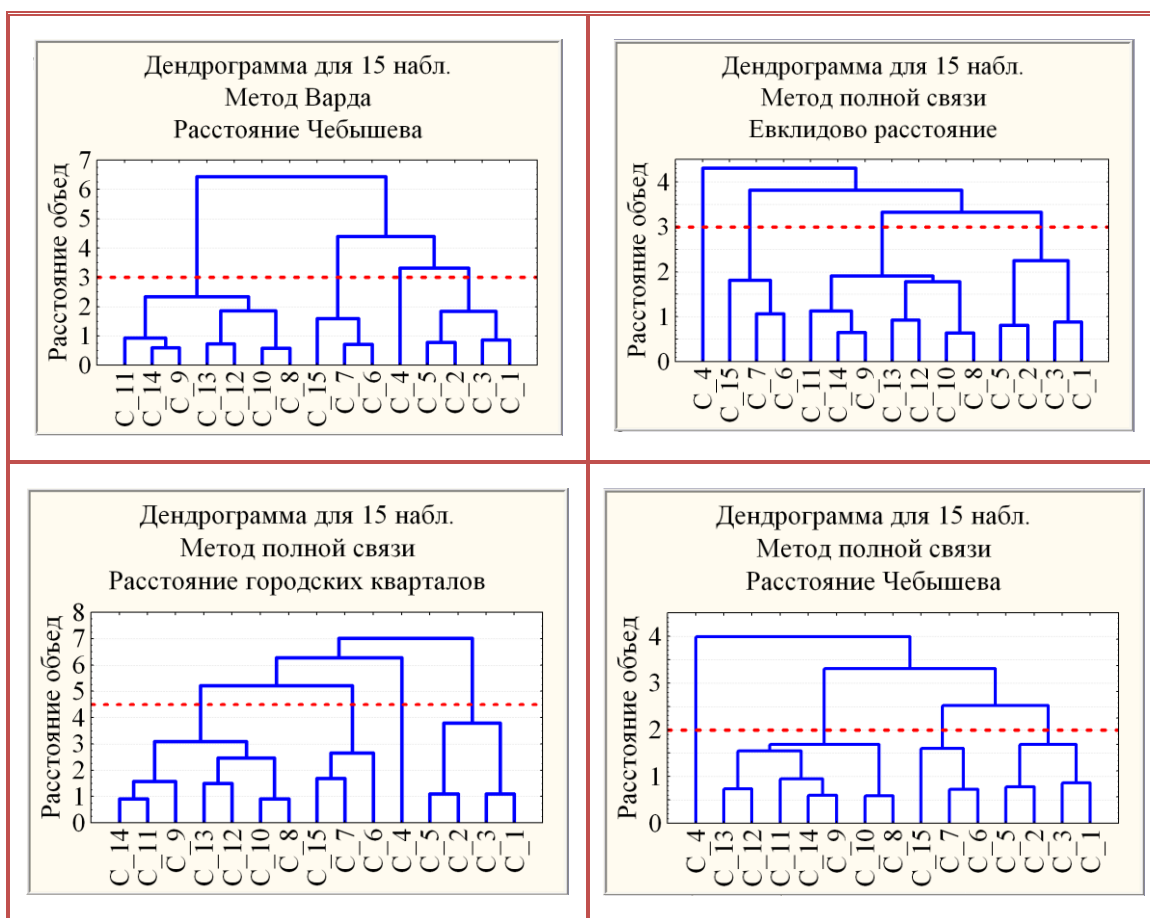



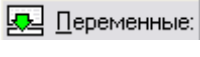

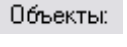

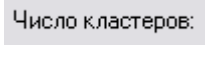
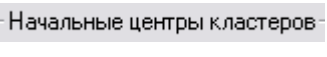


Рис. 3.221.  результатов иерархической классификации наблюдений Ч*І

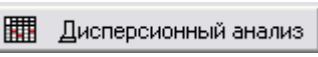
Таким образом, в построенном 3-х мерном факторном пространстве $\{\Phi_1, \Phi_2, \Phi_3\}$ проведена кластеризация наблюдений Ч*І. Методом древовидной кластеризации построено иерархическое дерево (рис.

3.221), стабильно делящее 15 наблюдений на 4 кластера: K1: {C8–C14}; K2: {C1–C3, C5}; K3: {C6, C7, C15} K4: {C4}.

Применим наряду с методом древовидной кластеризации также метод *K*-средних, проводящий классификацию объектов (наблюдений) по заданному количеству кластеров. При этом алгоритм метода *K*-средних, перемещая объекты в разные кластеры с целью минимизации изменчивости внутри кластеров и максимизации изменчивости между кластерами, оценивает качество кластеризации наблюдений по каждому фактору посредством параметрического дисперсионного анализа.

Шаг 2. В окне  (см. рис. 3.201) выберите метод  и щелкните . В открывшемся окне нажмите кнопку , выберите все факторные переменные аналогично рис. 3.201 и нажмите . В окне  выделите , в окне  выделите 4 в соответствии с рис. 3.221, а в окне  – переберите разные варианты опции, позволяющие выбрать способ вычисления начального расположения центров кластеризации, от которого до некоторой степени зависят результаты метода *K*-средних, если имеется много небольших кластеров, которые ясно различимы.

Рассмотрим результаты кластеризации наблюдений методом *K*-средних аналогично рис. 3.203.


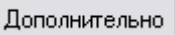
Шаг 3. Нажмите прежде всего кнопку  для построения стандартной таблицы результатов дисперсионного анализа кластеризации наблюдений по каждому фактору (рис. 3.222).

перемен.	Дисперсионный анализ (Ч*І)					
	Между SS	сс	Внутри SS	сс	F	значим. р
Ф1	8,771	3	5,229	11	6,151	0,01036
Ф2	9,283	3	4,717	11	7,216	0,00601
Ф3	12,386	3	1,614	11	28,138	0,00002

Рис. 3. 222. Результаты дисперсионного анализа кластеризации наблюдений Ч*І по факторам


Согласно рис. 222, значения *p*-уровней определяют статистически значимым (на уровне значимости $0,005 < p < 0,050$) различия между

кластерами для факторов Φ_1 и Φ_2 , высоко значимым (на уровне значимости $p \approx 0,00002 < 0,0005$) различия между кластерами для фактора Φ_3 . Таким образом, в отличие от $T \times I$ (рис. 3. 204) в данном случае выборки наблюдений $Ч \times I$ различия между кластерами являются значимыми (на разном уровне) для всех факторов.

Шаг 4. Нажмите кнопку  Средние кластеров и евклидовы расстояния в режиме  Дополнительно для вывода на экран таблицы результатов со средними для каждого кластера по каждому фактору (рис. 3.223) аналогично рис. 3.205.

перемен.	Средн.класт. (O_M)			
	Кластер Но. 1	Кластер Но. 2	Кластер Но. 3	Кластер Но. 4
Φ_1	0,672	-0,046	-1,332	-0,526
Φ_2	-0,521	1,230	-0,041	-1,151
Φ_3	0,366	-0,539	0,823	-2,873

Рис. 3.223. Кластерные средние наблюдений $Ч \times I$ в факторном пространстве

Шаг 4'. Нажмите кнопку  График средних для вывода на экран графиков факторных средних для каждого кластера (рис. 3.224) аналогично рис. 3.206.

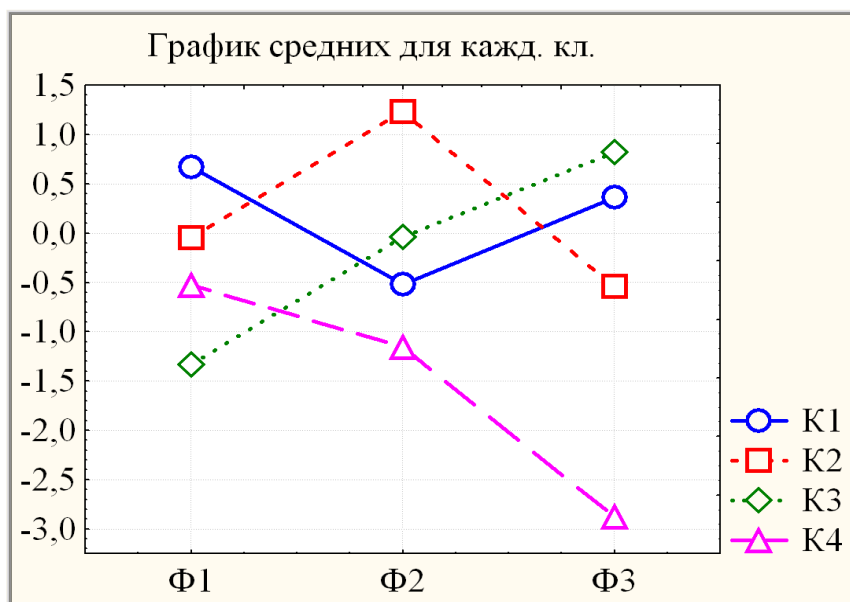


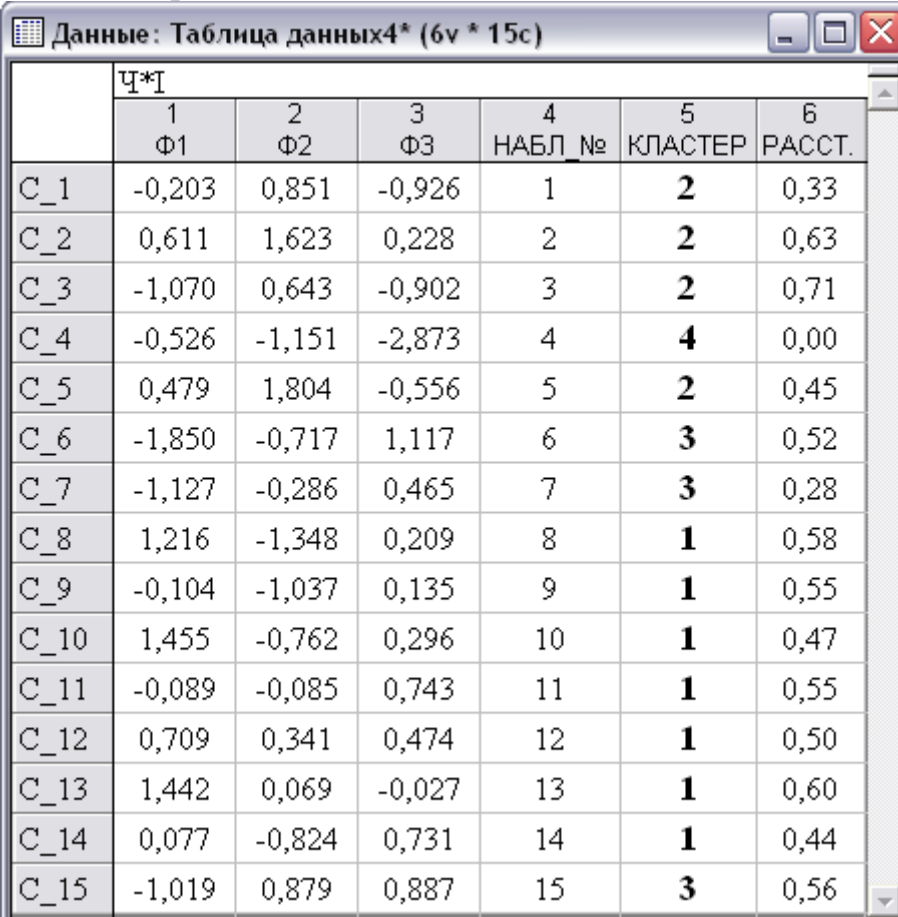


Рис. 3. 224. Линейные графики факторных средних наблюдений $Ч \times I$ для каждого кластера

Заметим, что с помощью  Элементы кластеров и расстояния (см. рис. 3.203) можно вывести на экран списки элементов (наблюдений) каждого кластера, где для каждого наблюдения вычислено также его расстояние до центра кластера. Полученные результаты кластеризации наблюдений методом *K*-средних являются устойчивыми по отношению к вариации опций в окне Начальные центры кластеров и в полной мере соответствуют иерархическому дереву (рис. 3.221), построенному методом древовидной кластеризации и стабильно делящему 7 наблюдений на 4 кластера (K1: {C8–C14}; K2: {C1–C3, C5}; K3: {C6, C7, C15} K4: {C4}).

Шаг 5. Нажмите  Сохранить классификацию и расстояния для вывода на экран данных, содержащий информацию о классификации и расстояниях наблюдений до центра соответствующего кластера, в дополнение к факторным переменным, выбранным из исходного файла данных (рис. 3.225) аналогично рис. 3.207.








	Ч*І					
	1 Ф1	2 Ф2	3 Ф3	4 НАБЛ. №	5 КЛАСТЕР	6 РАССТ.
C_1	-0,203	0,851	-0,926	1	2	0,33
C_2	0,611	1,623	0,228	2	2	0,63
C_3	-1,070	0,643	-0,902	3	2	0,71
C_4	-0,526	-1,151	-2,873	4	4	0,00
C_5	0,479	1,804	-0,556	5	2	0,45
C_6	-1,850	-0,717	1,117	6	3	0,52
C_7	-1,127	-0,286	0,465	7	3	0,28
C_8	1,216	-1,348	0,209	8	1	0,58
C_9	-0,104	-1,037	0,135	9	1	0,55
C_10	1,455	-0,762	0,296	10	1	0,47
C_11	-0,089	-0,085	0,743	11	1	0,55
C_12	0,709	0,341	0,474	12	1	0,50
C_13	1,442	0,069	-0,027	13	1	0,60
C_14	0,077	-0,824	0,731	14	1	0,44
C_15	-1,019	0,879	0,887	15	3	0,56


Рис. 3.225. Результаты факторной классификации наблюдений Ч*І

Заметим, что, согласно рис. 3. 221 и 3. 225, возможна более детальная кластеризация наблюдений Ч*І, при которой, например, кластер


К2 дробится на более однородные К21: {С1, С3} и К22: {С2, С5}, а от кластера К3 отделяется {С15} и т.д.


Так как значения одномерного F -критерия для всех трех факторов – значимы (рис. 3.222), то далее можно использовать метод множественных сравнений для выявления тех групп кластеров для каждого фактора, которые вызвали данный эффект и максимально отличаются от других, а также можно выделить однородные (различающиеся незначимо, то есть $p > 0,10$) иногда пересекающиеся группы кластеров.

Шаг 8. Запустите однофакторный  Дисперсионный анализ (ДА) аналогично рис. 3.208 и выберите в качестве зависимых все факторные переменные (Ф1-Ф3 для Ч*І), а в качестве категориальной КЛАСТЕР, закажите все коды (1-4) аналогично рис. 3.208. Затем в окне  Результаты анализа (см. рис.3.66) в режиме нажмите  Одномерн. результаты для просмотра результатов дисперсионного анализа (рис. 3.222). Далее в окне  Результаты анализа (рис. 3.66) нажмите кнопку и в открывшемся окне выберите режим , в котором представлены различные апостериорные критерии (рис. 3.209). Выберите, например, критерий  Фишера НЗР с опциями отображения Значимые различия и Однородные группы: (рис. 3. 226) аналогично рис. 3. 209.



НЗР крит.; перем. Ф1 (О_М) Однород гр, alpha = ,05000 Ошибка: Межгр. MS = ,4753				НЗР крит.; перем. Ф2 (О_М) Однород гр, alpha = ,05000 Ошибка: Межгр. MS = ,4288				НЗР крит.; перем. Ф3 (О_М) Однород гр, alpha = ,05000 Ошибка: Межгр. MS = ,1467				
	Ф1	1	2		Ф2	1	2		Ф3	1	2	3
	Среднее				Среднее				Среднее			
К3	-1,332		****	К4	-1,151	****		К4	-2,873		***	
К4	-0,526	****	****	К1	-0,521	****		К2	-0,539			***
К2	-0,046	****		К3	-0,041	****		К1	0,366	***		
К1	0,672	****		К2	1,230		****	К3	0,823	***		

Рис. 3.226. Результаты апостериорного критерия множественных сравнений средних  Фишера НЗР наблюдений Ч*І

Согласно рис. 3.226, для каждого фактора по апостериорному критерию  Фишера НЗР выделены следующие (упорядоченные по возрастанию среднего) однородные группы кластеров:

Ф1: {К3, К4} и {К4, К2, К1} так, что К3 ($m_{K3} \approx -1,332$) отличается от К2 ($m_{K2} \approx -0,046$) статистически значимо на уровне значимости $p \approx 0,033$.

Ф2: {К4, К1, К3} и {К2}, так, что К3 ($m_{K3} \approx -0,041$) отличается от К2 ($m_{K2} \approx 1,230$) статистически значимо на уровне значимости $p \approx 0,027$.

Ф3: {К4}, {К2} и {К1, К3} так, что К4 ($m_{K4} \approx -2,873$) отличается от К2 ($m_{K2} \approx -0,539$) высоко значимо на уровне значимости $p \approx 0,0002$, а К2 ($m_{K2} \approx -0,539$) отличается от К1 ($m_{K1} \approx 0,366$) сильно значимо на уровне значимости $p \approx 0,003$.

Заметим, что критерий НЗР дает самые значимые результаты по сравнению с другими апостериорными критериями.





Результаты кластерного анализа наблюдений Ч*Г по совокупности факторов с учетом результатов сравнений кластерных средних для каждого фактора (рис. 3.222-226) позволяют провести классификацию наблюдений в порядковой шкале стандартизированных измерений, полагая, например, в качестве уровня «Средний» – стандартизированный интервал (-1,0; +1,0) наблюдений, «Выше среднего» – (+1,0; +2,5) наблюдений, «Ниже среднего» – (-2,5; -1,0). Аномально низкие значения (< -2,5) определяют уровень «Аутсайдер».

Таблица 3.33

Классификация наблюдений Ч*Г по совокупности факторов в порядковой шкале стандартизированных измерений.

Кластер	Уровень кластера на фоне среднего по фактору			Состав кластера
	Ф1 (Na – Cr)	Ф2 (Fe – Ba – Sr)	Ф3 (Ca)	
К1	Средний	Средний	Средний	С_8–14
К2	Средний	Выше среднего	Средний	С_1–3, 5
К3	Ниже среднего	Средний	Средний	С_6, 7, 15
К4	Средний	Ниже среднего	Аутсайдер	С_4

С учетом малочисленности состава кластеров для достоверности полученных результатов параметрического дисперсионного анализа обратимся к непараметрическому однофакторному дисперсионному анализу на основе критерия Краскела-Уоллиса и медианного теста.

Шаг 9. Запустите  Непараметрическая статистика аналогично рис. 3.210, затем выберите  Сравнение нескольких независимых групп, закажите переменные Ч*І стандартным образом аналогично рис. 3.210 и щелкните по кнопкам  Да Кр.Краскела-Уоллиса и медианный тест и  Сравнения средних рангов для всех групп (рис. 3.227).

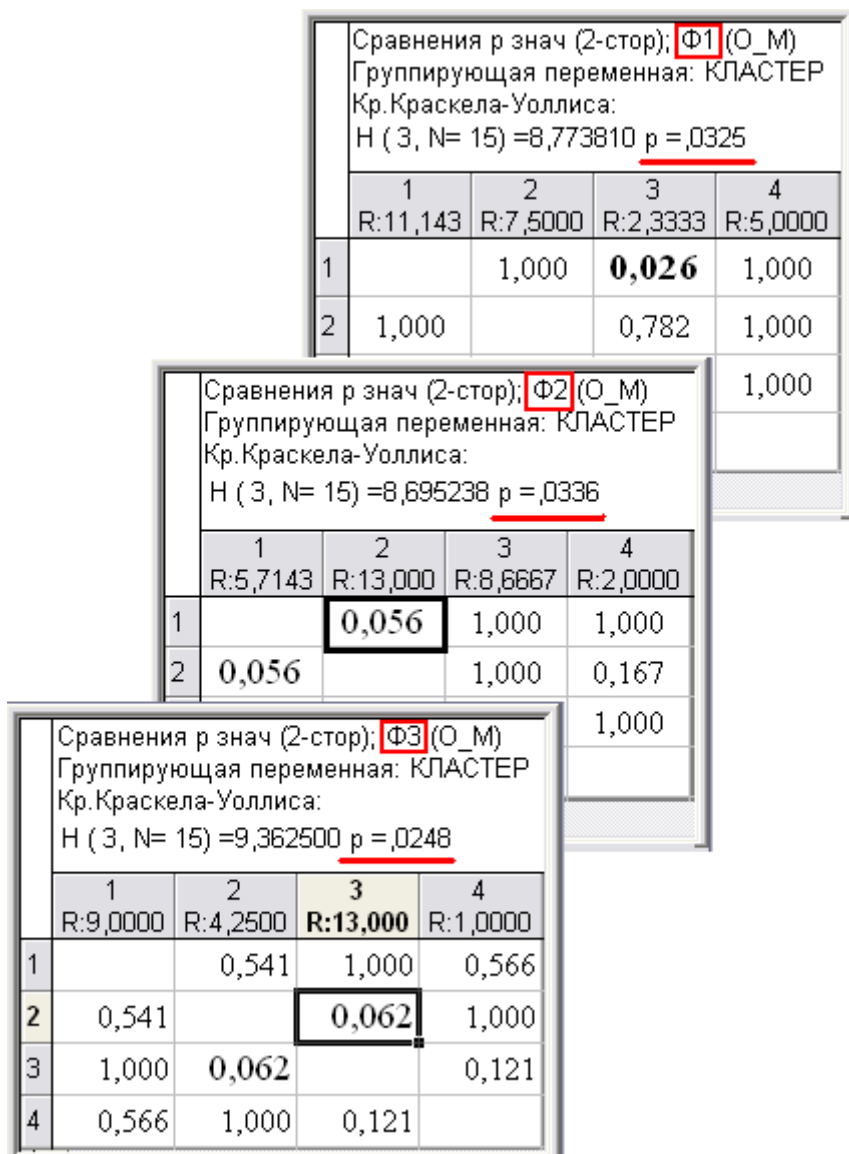


Рис. 3.227. Результаты рангового дисперсионного анализа кластеризации наблюдений Ч*І на основе критерия Краскела-Уоллиса

По совокупности критерия Краскела-Уоллиса (рис. 3.227) и медианного теста 4 кластера К1-К4 различаются статистически значимо (на уровне значимости $0,005 < p < 0,050$) по каждому фактору. Таким образом, непараметрический дисперсионный анализ лишь слегка «смягчает»

результаты параметрического дисперсионного анализа о значимости кластеризации наблюдений выборки Ч*І (рис. 3.222). С учетом этого обстоятельства в рамках параметрического подхода (сравнение кластерных средних) результаты классификацию наблюдений Ч*І в порядковой шкале (табл. 3.33) следует «смягчить», заменяя уровень «Аутсайдер» на «Ниже среднего» для моно – кластера К4 по фактору Ф3.

ВЫВОДЫ

✓ Согласно табл. 3.23 - 24, по совокупности данных трех критериев нормально распределенными одновременно в двух городах можно считать содержания микроэлементов **Na_{IS}, Fe, Sr, Ba, Cr**, а логнормально распределенным – содержания химических элементов **Na_{ИН}**. В данном случае логнормально распределенным – содержания химических элементов **Na_{ИН}**, определенного методом инструментального нейтронно-активационного анализа, можно объяснить тем, что для макроэлементов данный метод имеет существенный разброс значений. Его преимущества определять микроэлементы, что очень важно в геохимии редких и редкоземельных элементов.

✓ Как следует из табл.3.25, к неоднородным (различия в средних – значимые) по областям следует отнести выборки **Na_{ИН}, Ca_{ИН}, Fe_{IS}, Sr, Ba** и **Cr_{ИН}**.

✓ Согласно табл.3.26 (по Т) высоко значимо различаются по методам ИНАА и ISP содержания химических элементов Na, Ca и Cr, не значимо – Fe, Sr и Ba, а по табл.3.27 (по Ч) высоко значимо различаются по методам ИНАА и ISP содержания химических элементов Na, Ca, Fe, Sr и Ba, не значимо – Cr. В данном случае также существенна роль двух методов, что и нужно учитывать при аналитических исследованиях. Для атомно-эмиссионного анализа с индуктивно связанной плазмой (ISP) характерно более представительная навеска пробы, которая может измеряться несколькими граммами, тогда как для инструментального нейтронно-активационного анализа (ИНАА) – вес пробы до 100 мг. Все это необходимо учитывать при постановке задач для исследования.

✓ Согласно табл.3.28, результаты содержаний химического элемента в накипи по методам ИНАА и ISP можно считать не коррелированными для Na в (Ч), Ca и Fe в (Т), Cr в (Т) и (Ч); сильно коррелированными для Na в (Т), Ca и Fe в (Ч), Sr и Ba в (Т) и (Ч).

✓ Корреляционные матрицы (табл.3.29-32) и дендрограммы (рис.3.38-41) геохимического спектра микроэлементов в накипи Т и Ч по методам ИНАА и ISP выделяют одну общую по населенным пунк-

там и методам значимую последовательную ассоциацию Ba-Sr-Fe, а также Cr-Na по населенным пунктам для метода ISP. Ассоциация Ba-Sr-Fe элементов имеет общую с геохимических позиций основу. Макроэлементы, специфика накипи как для Томского региона, так и для Челябинского, которая характеризует природную составляющую в данных объектах исследования.

✓ Метод канонических корреляций построены две статистически значимые пары канонических переменных, соответствующих межгрупповым корреляционно значимым микроассоциациям Ca-Sr и Ba-Sr-Fe, объясняющим $\approx 55\%$ дисперсии переменных как в левом (Ч*И), так и в правом (Ч*И) множествах, что свидетельствует о достаточно сильной корреляционной зависимости между переменными двух множеств Ч*И и Ч*І.

✓ В рамках факторного анализа методом главных компонент выделена совокупность первых 3-х собственных значений, объясняющая $\approx 95,43\%$ от общей дисперсии, и построена 3-х факторная модель микроэлементов Т*І: Ф1 (Na – Cr), Ф2 (Ca) и Ф3 (Ba – Sr – Fe).

✓ На базе построенной факторной модели содержаний микроэлементов Т*І проведена кластеризация наблюдений Т*І и построена их 4-х кластерная модель, делящее 7 наблюдений на 4 кластера (К1: С1+С4; К2: С2+С7; К3: С3+С5 К4: С6). Аналогичным образом проведена кластеризация наблюдений Ч*І и построена их 4-х кластерная модель, делящее 15 наблюдений на 4 кластера (К1: С_8–14; К2: С_1–3, 5; К3: С_6, 7, 15; К4: С4). Качество проведенной кластеризация оценено критериями непараметрического и параметрического однофакторного дисперсионного анализа. В данном случае выделяются кластеры, которые учитывают многие специфические особенности проб накипи населенных пунктов. В тех случаях, когда ведется водоподготовка на ранних стадиях, четко фиксируются свои особенности, которые проявляются в более низких концентрациях элементов в накипи. Это хорошо проявляется в населенных пунктах Наумовка (С2) и Козюлино (С7). Несколько по иному фиксируется кластер К1, где отбор проб воды в населенных пунктах был произведен из верховодки. Вода не проходит промышленную водоподготовку и характеризуется присутствием высоких концентраций большого количества элементов техногенного характера. В данном случае опробование проводилось в частном секторе (С1 – с. Черная Речка и С4 – с. Орловка. Совсем другая ситуация в кластере К4, где пробы изучены в с. Моряковка. В данном случае особая геохимическая редкоземельная специализация населенного пункта, которая проявляется по многим природным объектам, в т.ч. и в снеге и почвах.

П₂. Содержание микроэлементов (мг/кг) в почве территорий гг. Междуреченска (М) и Томска (Т), полученное методом ЭСП.

Таблица 3.34

Раздельная составляющая П₂ (М).

	Mn_М	Ва_М	Сг_М	V_М	Ni_М	Со_М	Сu_М	Pb_М	Zn_М
Ед. Изм.	0,1%	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %
1	0,5	40	4	5	4	2,5	10	4	15
2	0,6	30	4	4	3	2	5	4	15
3	0,4	30	6	5	4	3,5	7	5	10
4	0,4	30	2	3	2	1	4	5	20
5	0,4	30	4	5	7	3	6	5	10
6	0,2	30	6	5	4	1	4	2	6
7	0,2	40	2	2	2	1	5	3	8
8	0,2	80	10	6	4	1,5	4	2	8
9	0,1	30	4	3	3	1	4	4	10
10	0,1	20	8	5	5	2	9	1	15
11	0,2	50	6	3	3	1	4	4	15
12	0,2	30	6	4	3	1,5	2	3	8
13	0,3	40	4	5	4	1,5	4	4	10
14	0,5	40	4	5	4	1,5	5	4	15
15	0,2	30	2	3	3	0,4	2	1	6
16	0,2	30	4	4	3	0,8	3	4	8
17	0,4	30	6	4	3	1,5	4	3	8
18	0,2	30	6	5	4	1,5	4	4	15
19	0,5	50	6	3	2	0,8	3	5	15
20	0,4	30	14	5	5	2	4	6	10
21	0,4	30	2	2	2	0,4	2	1	6
22	0,6	50	4	1	2	0,4	4	3	8
23	0,4	50	2	3	3	0,8	3	3	8
24	0,6	30	4	9	6	3	5	4	8
25	0,4	30	6	4	3	2	5	3	10
26	0,5	50	8	6	4	3	6	5	10
27	0,4	30	8	6	4	2,5	6	3	10
28	0,6	20	4	5	6	3	5	3	10
29	0,5	50	6	5	4	1,5	5	5	10
30	0,7	80	12	10	5	2,5	5	6	15

Продолжение табл. 3.34

	Mn _M	Ba _M	Cr _M	V _M	Ni _M	Co _M	Cu _M	Pb _M	Zn _M
31	0,4	60	10	8	5	1,5	5	4	8
32	0,2	30	4	3	3	0,4	4	1	8
33	0,3	50	10	6	3	0,8	3	4	6
34	0,4	30	4	2	2	0,8	4	2	6
35	0,7	70	8	6	2	0,8	3	5	8
36	0,7	80	10	6	4	1	3	5	6
37	0,6	30	6	4	4	2	5	3	8
38	0,4	60	10	6	3	0,8	3	3	4
39	0,8	30	8	5	4	2	5	4	10
40	0,6	60	8	5	5	2,5	6	6	20
41	0,3	40	6	3	3	1	4	2	8
42	0,4	30	8	5	7	3	4	3	15
43	0,4	30	6	4	4	1	6	4	15
44	0,1	30	4	3	3	0,8	5	4	10
45	0,1	20	6	4	4	2,5	5	2	8
46	0,3	50	8	6	2	0,4	2	2	6
47	0,2	20	6	3	3	1,5	4	2	8
48	0,3	50	4	7	3	3	6	4	10
49	0,6	50	6	7	4	1	4	3	4
50	0,7	60	6	7	3	2,5	7	4	10
51	0,8	70	6	7	3	2,5	6	3	8
52	0,3	60	10	7	4	0,8	3	4	4
53	0,5	40	10	7	3	1	6	3	4
54	0,3	30	6	7	3	2,5	7	2	8
55	0,3	50	6	6	3	2	8	5	15
56	0,3	50	6	4	3	1,5	5	5	10
57	0,4	30	8	6	3	3	5	2	10
58	0,2	30	6	6	2	1,5	4	2	10
59	0,3	60	8	8	4	0,4	2	4	6
60	0,4	40	4	8	3	3	7	4	15
61	0,6	60	10	10	5	1,5	3	3	8
62	0,6	50	6	8	3	3,5	8	6	15
63	0,5	40	10	8	4	1	4	4	20
64	0,2	30	8	6	3	2	5	2	8
65	0,7	40	10	7	4	4	9	5	15

Окончание табл. 3.34

	Mn _M	Ba _M	Cr _M	V _M	Ni _M	Co _M	Cu _M	Pb _M	Zn _M
66	0,8	70	8	7	3	2,5	10	6	10
67	0,8	40	10	8	4	4	9	3	15
68	0,2	30	8	7	4	2,5	7	3	10
69	0,3	50	6	8	5	2	6	5	6
70	0,5	40	4	4	1	1	6	4	10
71	0,5	40	6	6	3	2,5	7	3	8
72	0,1	20	2	1	1	0,8	7	1	8
73	0,3	40	8	10	5	1	4	3	6
74	0,8	80	12	6	3	1	3	4	8
75	0,7	70	4	7	4	1,5	5	4	6
76	0,1	40	6	7	3	2	6	3	8
77	0,1	30	4	4	3	0,4	3	3	6
78	0,2	40	8	7	3	2	6	3	8
79	0,7	60	10	5	4	3	5	2	15
80	0,3	60	8	5	4	3	7	2	10
81	0,3	40	6	6	3	1	5	4	6
82	0,3	50	4	6	3	0,4	3	3	4
83	0,5	40	4	1	2	0,4	8	4	8
84	0,4	30	6	6	3	2	6	2	10
85	0,1	30	2	4	2	0,4	4	1	8
86	0,5	40	8	8	4	1,5	4	3	8
87	0,3	50	6	8	3	0,4	3	3	4
88	0,5	40	8	8	3	2	6	4	10
89	0,4	30	8	6	2	2	6	2	10
90	0,5	20	10	9	6	1,5	4	3	8
91	0,5	30	10	8	4	3	6	2	10
92	0,6	30	10	9	5	2	5	4	8
93	0,2	30	12	9	5	2,5	8	4	8
94	0,1	20	8	7	3	2	6	2	10
95	0,4	30	8	4	3	2	4	2	10
96	0,5	40	10	9	5	1,5	5	4	10
97	0,1	30	2	5	3	0,4	2	1	4
98	0,1	30	6	4	4	1,5	4	2	8
99	0,1	20	8	9	4	3,5	7	2	10
100	0,3	40	8	7	3	2,5	7	4	20

Таблица 3.35

Раздельная составляющая Π_2 (Т).

	Mn_T	Ba_T	Cr_T	V_T	Ni_T	Co_T	Cu_T	Pb_T	Zn_T
Ед. Изм.	0,1%	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %
1	0,7	50	12	10	4	2,5	6	3	5
2	0,5	50	12	8	4	2,5	6	3	4
3	0,6	30	17	10	5	3,5	6	4	5
4	0,4	70	17	7	4	2,5	6	3	4
5	0,5	60	17	9	5	3	6	4	5
6	0,7	70	17	8	4	1,5	6	4	6
7	0,3	40	8	7	3	1,5	4	4	5
8	0,1	40	8	3	3	1,5	10	4	4
9	0,4	50	12	9	7	4	9	5	6
10	0,2	30	6	7	4	3	8	5	5
11	0,3	50	10	8	3	2,5	8	2	5
12	0,2	40	12	5	2	1	9	4	6
13	0,4	60	6	3	3	1,5	6	2	5
14	0,5	40	10	9	2	3	8	4	8
15	0,3	40	10	8	3	2,5	8	4	6
16	0,5	50	10	3	2	1,5	8	3	4
17	0,1	30	6	3	2	1,5	10	4	6
18	0,3	50	12	8	3	2	7	2	5
19	0,3	50	17	8	3	2,5	7	3	4
20	0,3	30	8	7	4	4	9	3	6
21	0,3	40	12	4	2	1	9	4	5
22	0,5	50	10	7	5	2,5	7	3	6
23	0,1	30	6	5	3	1	7	3	5
24	0,2	40	6	7	3	2	7	2	5
25	0,3	40	12	9	4	3,5	7	2	6
26	0,2	30	6	5	2	1	7	3	3
27	0,5	40	8	3	2	1,5	7	2	5
28	0,5	40	15	7	4	3,5	7	2	6
29	0,2	40	10	8	2	2,5	7	4	10
30	0,3	40	20	7	3	2	6	2	4
31	0,6	40	20	7	5	1,5	6	3	6
32	0,4	60	4	2	3	0,8	8	6	10

Продолжение табл. 3.35

	Mn_T	Ba_T	Cr_T	V_T	Ni_T	Co_T	Cu_T	Pb_T	Zn_T
33	0,5	40	20	5	6	1	6	3	6
34	0,6	40	15	7	5	2	6	2	5
35	0,5	20	15	5	6	1,5	8	5	8
36	0,6	40	15	7	5	2	6	2	5
37	0,1	20	2	1	3	0,4	5	2	10
38	0,5	30	6	5	4	1	8	3	15
39	0,4	20	10	5	3	1,5	5	1	4
40	0,6	30	8	6	3	1,5	5	2	8
41	0,5	40	8	3	2	1,5	7	5	5
42	0,4	40	12	5	2	1,5	5	2	6
43	0,4	30	8	7	3	2	7	2	5
44	0,5	40	8	6	3	1,5	7	3	8
45	0,7	40	12	5	3	1,5	8	4	10
46	0,7	40	12	6	3	1,5	2	3	6
47	0,8	40	10	9	4	1,5	6	4	10
48	0,2	40	12	3	2	0,8	3	2	5
49	0,2	30	2	4	3	0,8	5	5	8
50	0,5	60	17	6	4	1,5	4	4	10
51	0,4	30	6	7	5	3	4	3	8
52	0,5	50	15	8	6	3	4	4	15
53	0,6	30	8	5	4	2,5	10	6	15
54	0,4	70	10	4	3	0,8	6	4	20
55	0,7	70	12	8	4	1,5	6	2	10
56	0,3	20	20	8	4	1,5	6	1	6
57	0,1	20	10	6	3	1,5	3	1	4
58	0,2	30	15	4	3	0,4	2	1	4
59	0,3	60	20	6	3	0,8	3	1	6
60	0,3	60	8	6	2	0,4	3	1	5
61	0,7	80	4	4	2	0,8	6	3	6
62	0,2	30	10	4	2	0,8	4	2	4
63	0,3	30	8	6	3	1,5	6	4	10
64	0,8	30	8	7	4	0,4	3	1	6
65	0,6	30	4	6	2	2	7	4	20
66	0,3	40	6	4	3	1	5	2	10
67	0,7	70	8	6	3	2	5	2	10

Окончание табл. 3.35

	Mn _T	Ba _T	Cr _T	V _T	Ni _T	Co _T	Cu _T	Pb _T	Zn _T
68	0,4	30	10	8	3	1,5	5	1	6
69	0,1	30	8	6	3	2	5	2	5
70	0,4	30	4	5	4	2,5	9	4	15
71	0,6	50	8	6	3	2	5	1	8
72	0,4	30	4	4	3	0,4	3	1	3
73	0,7	70	17	9	4	3	2	2	3
74	0,3	20	20	5	3	1,5	5	1	6
75	0,4	30	15	10	5	3	5	2	4
76	0,2	20	15	9	3	2	4	2	4
77	0,6	20	15	9	3	3	4	3	4
78	0,8	30	15	9	3	2	5	2	3
79	0,1	20	6	5	2	0,4	2	2	3
80	0,8	30	10	6	3	0,8	4	2	4
81	0,4	30	17	7	3	2	4	3	3
82	0,5	70	10	7	3	1,5	3	3	8
83	0,4	30	10	8	4	3	4	3	5
84	0,6	30	8	9	4	2	4	3	5
85	0,4	20	10	7	4	2	6	6	15
86	0,7	80	12	10	4	0,8	4	3	8
87	0,4	20	6	1	3	0,4	3	3	6
88	0,3	60	17	8	4	1	5	5	8
89	0,6	50	12	10	4	1,5	3	3	5
90	0,4	30	12	6	3	0,8	3	4	4
91	0,5	80	10	6	3	1,5	4	2	8
92	0,4	30	15	9	3	1,5	5	4	8
93	0,6	60	15	10	4	1,5	6	4	8
94	0,4	60	12	10	4	1,5	5	6	10
95	0,2	30	10	1	3	0,3	2	2	3
96	0,2	30	15	6	3	0,4	2	2	10
97	0,2	30	6	4	1	1	4	1	4
98	0,2	30	6	4	2	1	5	2	5
99	0,6	30	17	6	2	1,5	4	3	5
100	0,6	20	15	4	2	1	5	1	4

Таблица 3.36

Категоризированная составляющая П₂

	Город	Mn	Ba	Cr	V	Ni	Co	Cu	Pb	Zn
Ед. изм.		0,1%	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %	10 ⁻³ %
1	М	0,5	40	4	5	4	2,5	10	4	15
2	М	0,6	30	4	4	3	2	5	4	15
3	М	0,4	30	6	5	4	3,5	7	5	10
4	М	0,4	30	2	3	2	1	4	5	20
5	М	0,4	30	4	5	7	3	6	5	10
...
96	М	0,5	40	10	9	5	1,5	5	4	10
97	М	0,1	30	2	5	3	0,4	2	1	4
98	М	0,1	30	6	4	4	1,5	4	2	8
99	М	0,1	20	8	9	4	3,5	7	2	10
100	М	0,3	40	8	7	3	2,5	7	4	20
101	Т	0,7	50	12	10	4	2,5	6	3	5
102	Т	0,5	50	12	8	4	2,5	6	3	4
103	Т	0,6	30	17	10	5	3,5	6	4	5
104	Т	0,4	70	17	7	4	2,5	6	3	4
105	Т	0,5	60	17	9	5	3	6	4	5
...
196	Т	0,2	30	15	6	3	0,4	2	2	10
197	Т	0,2	30	6	4	1	1	4	1	4
198	Т	0,2	30	6	4	2	1	5	2	5
199	Т	0,6	30	17	6	2	1,5	4	3	5
200	Т	0,6	20	15	4	2	1	5	1	4

П₃. Содержание микроэлементов (мг/кг) в почве (П), снеговом (С) осадке и золе (З) растений территории г. Стрежевого.

Таблица 3.37

Раздельная составляющая П₃ (С)

С	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
1	100	40	40	8	4	20	3	100	15	300	400
2	100	40	60	10	6	15	4	100	10	300	300
3	40	60	40	8	6	10	2	200	10	300	300
4	80	60	60	10	3	150	3	300	10	200	300
5	100	150	80	40	8	100	3	600	30	300	400
6	100	100	80	30	8	150	6	400	60	300	600
7	80	150	150	30	6	150	6	600	15	200	300
8	200	150	60	40	8	60	4	400	80	300	800
9	100	60	80	20	4	40	3	300	40	300	400
10	100	80	80	10	3	20	3	300	60	200	300
11	150	60	60	10	3	15	4	300	30	200	300
12	40	40	40	15	3	15	2	300	30	200	300

Таблица 3.38

Раздельная составляющая П₃ (З)

З	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
1	3,2	176	15	18	1,6	16,5	3	420	3	240	140
2	5,4	176	15	12	1,6	11	2	560	3	240	210
3	6	117,3	20	9	0,6	11	2	420	4	240	210
4	1,4	117,3	10	4,8	1,2	8,8	3	280	1	480	280
5	1,8	117,3	20	24	3	16,5	3	560	4	480	280

Таблица 3.39

Раздельная составляющая П₃ (П)

П	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
1	15	60	150	30	10	15	3	400	40	100	300
2	16	60	150	30	8	15	3	300	30	100	200
3	25	80	150	30	8	15	2	400	30	100	300
4	12	60	100	30	10	15	2	400	30	100	400
5	27	60	150	20	8	10	2	400	30	100	300
6	18	40	100	6	3	6	2	200	10	100	100
7	24	150	200	30	10	20	2	400	30	100	300
8	13	40	150	20	10	10	2	300	40	200	400
9	11	40	200	30	15	10	3	300	40	200	400
10	30	40	200	15	3	10	2	300	40	100	300
11	12	60	150	40	15	15	3	400	40	200	1000
12	13	40	300	15	3	20	2	300	30	100	300
13	9	40	60	40	15	15	3	600	30	100	400
14	11	60	100	30	10	20	3	300	30	100	200
15	16	80	300	60	15	20	4	400	40	100	400
16	15	40	150	40	15	20	3	600	40	200	600
17	10	40	300	10	3	10	2	300	20	100	100
18	59	80	150	60	20	30	3	600	40	200	400
19	11	40	200	30	15	15	4	400	30	200	400
20	20	60	100	20	10	15	2	400	30	200	400
21	30	150	200	30	6	60	3	600	30	200	300
22	33	100	100	30	10	30	3	600	30	100	300
23	24	60	150	40	15	20	3	600	40	200	400
24	35	100	200	30	10	20	3	400	30	100	400
25	11	40	150	30	15	20	3	400	40	100	400
26	30	150	100	40	20	30	3	600	60	200	600
27	15	40	200	30	10	15	3	400	30	100	300
28	47	100	150	30	4	15	3	400	30	100	300
29	20	60	150	20	4	10	3	300	30	100	300
30	40	60	150	30	6	20	3	300	40	200	300
31	71	300	100	30	6	20	2	600	30	100	300
32	11	40	100	20	4	8	2	300	20	100	300
33	11	60	150	30	6	30	3	400	30	100	200

Таблица 3.40

Категоризированная составляющая П₃

	Среда	Pb	Zn	Cr	Ni	Co	Cu	Mo	Mn	V	Sr	Ba
1	Почва	15	60	150	30	10	15	3	400	40	100	300
2	Почва	16	60	150	30	8	15	3	300	30	100	200
3	Почва	25	80	150	30	8	15	2	400	30	100	300
4	Почва	12	60	100	30	10	15	2	400	30	100	400
5	Почва	27	60	150	20	8	10	2	400	30	100	300
6	Почва	18	40	100	6	3	6	2	200	10	100	100
7	Почва	24	150	200	30	10	20	2	400	30	100	300
8	Почва	13	40	150	20	10	10	2	300	40	200	400
9	Почва	11	40	200	30	15	10	3	300	40	200	400
...
25	Почва	11	40	150	30	15	20	3	400	40	100	400
26	Почва	30	150	100	40	20	30	3	600	60	200	600
27	Почва	15	40	200	30	10	15	3	400	30	100	300
28	Почва	47	100	150	30	4	15	3	400	30	100	300
29	Почва	20	60	150	20	4	10	3	300	30	100	300
30	Почва	40	60	150	30	6	20	3	300	40	200	300
31	Почва	71	300	100	30	6	20	2	600	30	100	300
32	Почва	11	40	100	20	4	8	2	300	20	100	300
33	Почва	11	60	150	30	6	30	3	400	30	100	200
<hr/>												
34	Снег	100	40	40	8	4	20	3	100	15	300	400
35	Снег	100	40	60	10	6	15	4	100	10	300	300
36	Снег	40	60	40	8	6	10	2	200	10	300	300
37	Снег	80	60	60	10	3	150	3	300	10	200	300
38	Снег	100	150	80	40	8	100	3	600	30	300	400
39	Снег	100	100	80	30	8	150	6	400	60	300	600
40	Снег	80	150	150	30	6	150	6	600	15	200	300
41	Снег	200	150	60	40	8	60	4	400	80	300	800
42	Снег	100	60	80	20	4	40	3	300	40	300	400
43	Снег	100	80	80	10	3	20	3	300	60	200	300
44	Снег	150	60	60	10	3	15	4	300	30	200	300
45	Снег	40	40	40	15	3	15	2	300	30	200	300
<hr/>												
46	Зола	3,2	176	15	18	1,6	16,5	3	420	3	240	140
47	Зола	5,4	176	15	12	1,6	11	2	560	3	240	210
48	Зола	6	117,3	20	9	0,6	11	2	420	4	240	210
49	Зола	1,4	117,3	10	4,8	1,2	8,8	3	280	1	480	280
50	Зола	1,8	117,3	20	24	3	16,5	3	560	4	480	280

П₄. Категоризированное содержание микроэлементов в накипи территорий Томской (Т) и Челябинской (Ч) областей, полученное методами ИНАА и ISP (табл 5.10)

Таблица 3.41

	Обл	Метод	Na	Ca	Fe	Sr	Ba	Cr
1	Ч	ИНАА	0,170	29,0	0,89	0,296	0,121	25,3
2	Ч	ИНАА	0,130	48,6	0,87	0,499	0,117	20,8
3	Ч	ИНАА	0,090	29,8	0,66	0,284	0,093	18,1
4	Ч	ИНАА	0,060	5,6	0,26	0,068	0,007	21,5
5	Ч	ИНАА	0,070	35,8	1,20	0,371	0,149	23,1
6	Ч	ИНАА	0,070	54,4	0,52	0,301	0,046	25,7
7	Ч	ИНАА	0,060	42,5	0,52	0,306	0,081	24,4
8	Ч	ИНАА	0,100	49,9	0,74	0,270	0,031	26,8
9	Ч	ИНАА	0,080	53,8	0,35	0,160	0,038	29,0
10	Ч	ИНАА	0,090	50,0	0,84	0,422	0,037	21,5
11	Ч	ИНАА	0,060	52,7	0,46	0,344	0,042	22,1
12	Ч	ИНАА	0,090	54,8	0,47	0,397	0,056	22,2
13	Ч	ИНАА	0,080	49,9	0,48	0,330	0,045	19,6
14	Ч	ИНАА	0,070	52,9	0,45	0,239	0,051	21,6
15	Ч	ИНАА	0,060	52,4	1,30	0,360	0,044	24,8
16	Т	ИНАА	0,003	22,9	0,58	0,010	0,027	0,1
17	Т	ИНАА	0,010	29,1	0,56	0,073	0,014	0,3
18	Т	ИНАА	0,012	28,0	1,50	0,091	0,032	0,6
19	Т	ИНАА	0,003	23,9	0,21	0,048	0,024	1,3
20	Т	ИНАА	0,015	25,8	1,30	0,155	0,055	0,4
21	Т	ИНАА	0,003	29,4	0,75	0,052	0,008	0,1
22	Т	ИНАА	0,003	25,8	0,42	0,041	0,017	1,3
23	Т	ИНАА	0,017	23,7	0,11	0,042	0,020	0,4
24	Ч	ISP	1,830	25,5	0,40	0,217	0,010	27,0
25	Ч	ISP	0,890	36,9	0,44	0,361	0,014	16,0
26	Ч	ISP	2,480	21,7	0,27	0,194	0,011	38,0
27	Ч	ISP	2,330	3,6	0,02	0,029	0,001	41,0
28	Ч	ISP	1,050	34,5	0,70	0,281	0,012	19,0
29	Ч	ISP	2,690	37,5	0,11	0,203	0,002	46,0
30	Ч	ISP	2,400	34,1	0,05	0,173	0,009	35,0
31	Ч	ISP	0,630	37,3	0,02	0,196	0,000	11,0
32	Ч	ISP	1,630	36,8	0,03	0,096	0,006	25,0
33	Ч	ISP	0,620	40,5	0,10	0,202	0,005	3,0

Окончание табл. 3.41								
	Обл	Метод	Na	Ca	Fe	Sr	Ba	Cr
34	Ч	ISP	1,510	38,9	0,10	0,245	0,009	23,0
35	Ч	ISP	0,790	39,0	0,11	0,266	0,013	16,0
36	Ч	ISP	0,370	37,0	0,14	0,244	0,010	8,0
37	Ч	ISP	1,600	41,4	0,09	0,169	0,005	18,0
38	Ч	ISP	2,260	41,1	0,43	0,262	0,010	30,0
39	T	ISP	1,950	35,2	0,79	0,037	0,017	35,0
40	T	ISP	2,500	37,3	0,48	0,062	0,010	34,0
41	T	ISP	2,080	37,0	1,24	0,066	0,021	29,0
42	T	ISP	1,470	35,3	1,10	0,044	0,017	25,0
43	T	ISP	2,320	36,8	1,05	0,114	0,034	38,0
44	T	ISP	0,600	37,6	0,57	0,037	0,005	11,0
45	T	ISP	2,070	38,0	0,36	0,039	0,011	33,0
46	T	ISP	2,750	38,8	0,14		0,017	42,0