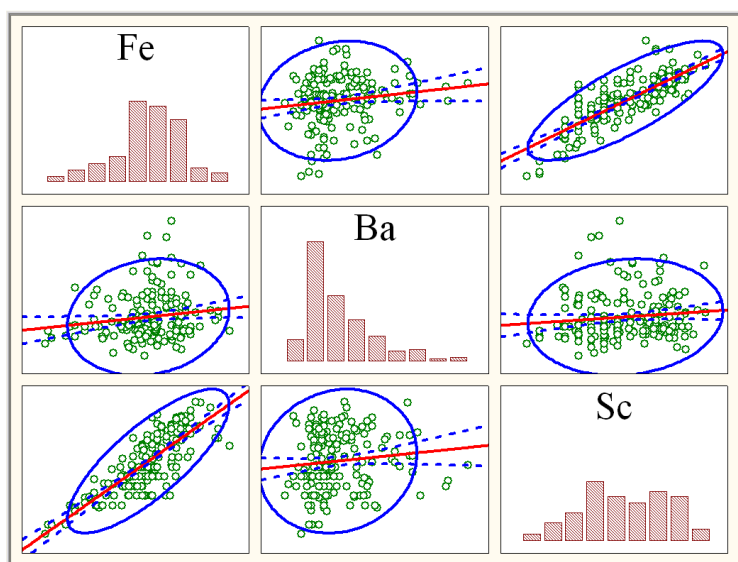


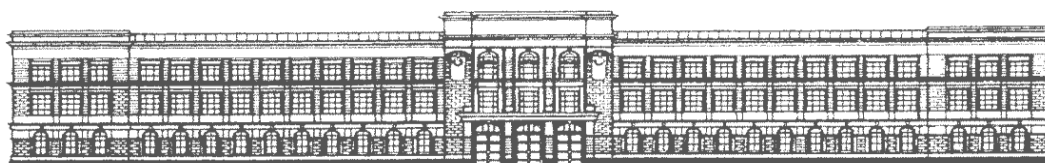


А.А. Михальчук, Е.Г. Язиков

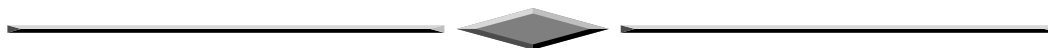
МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ



ЭКОЛОГО-ГЕОХИМИЧЕСКИХ ИЗМЕРЕНИЙ



**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**



А.А. Михальчук, Е.Г. Язиков

**МНОГОМЕРНЫЙ
СТАТИСТИЧЕСКИЙ АНАЛИЗ
ЭКОЛОГО-ГЕОХИМИЧЕСКИХ
ИЗМЕРЕНИЙ**

ЧАСТЬ II. КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Издательство ТПУ
Томск 2014

УДК 681.3 : 519.2

А.А. Михальчук, Е.Г. Язиков

Многомерный статистический анализ эколого-геохимических измерений. Часть II. Компьютерный практикум.

Учебное пособие. - Томск: Изд. ТПУ, 2014.- 150 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 II. КОМПЬЮТЕРНЫЙ ПРАКТИМ	4
2.1. Описательная статистика	5
2.1.1. Вычисление оценок числовых характеристик содержаний химических элементов.....	5
2.1.2. Построение диаграммы размаха.....	7
2.1.3. Построение гистограммы содержаний микроэлементов.....	9
2.2. Проверка статистических гипотез	15
2.2.1. Проверка гипотезы о законе распределения содержаний химических элементов.....	15
2.2.2. Оценка различия содержаний двух выборок.....	19
2.3. Компьютерный дисперсионный анализ	27
2.3.1. Однофакторный дисперсионный анализ.....	28
2.3.2. Двухфакторный дисперсионный анализ.....	39
2.4. Корреляционно – регрессионный анализ	47
2.4.1. Вычисление корреляционной матрицы ассоциации содержаний микроэлементов.....	47
2.4.2. Вычисление коэффициента корреляции Спирмена.....	52
2.4.3. Построение диаграммы рассеяния и линии регрессии.....	54
2.5. Канонический анализ двух ассоциации микроэлементов...	66
2.6. Кластерный анализ ассоциации микроэлементов.....	82
2.7. Факторный анализ ассоциации микроэлементов.....	90
2.8 Кластерный анализ наблюдений в факторной модели ассоциации микроэлементов	102
2.9. Средства статистического анализа данных в системе Excel	118
2.10. Модульный анализ данных в системе STATISTICA	138
ПРИЛОЖЕНИЕ	147
П₁ Содержание микроэлементов в почве территории Томского региона, полученное методами ИНАА и ISP	147

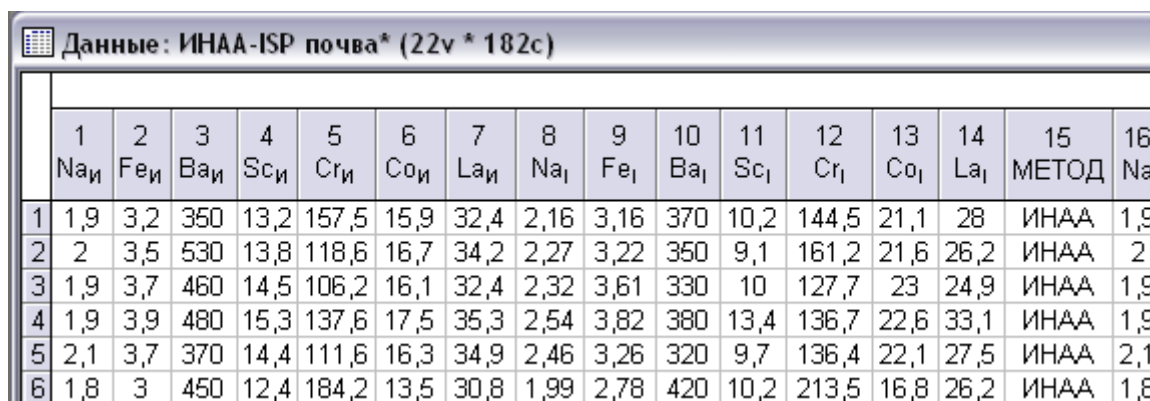
ЧАСТЬ II. КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Многомерный статистический анализ содержаний микроэлементов, полученных при проведении эколого-геохимических исследований разными методами, предполагает **поэтапно**:

- I. Вычисление оценок числовых характеристик содержаний.
- II. Проведение оценки различия содержаний по двум методам:
 1. проверка нормальности распределений содержаний,
 2. применение параметрических или непараметрических (ранговых) критериев различия (неоднородности) выборок.
- III. Проведение оценки степени зависимости между микроэлементами на основе корреляционно-регрессионного анализа:
 1. вычисление корреляционной матрицы двух выборок и оценку значимости коэффициента корреляции,
 2. построение диаграммы рассеяния и уравнения регрессии для наиболее значимых корреляционных связей в ассоциациях.
- IV. Проведение канонического анализа двух ассоциаций микроэлементов.
- V. Проведение оценки степени сходства результатов, полученных разными методами, на основе метода иерархической группировки (кластерного анализа).
- VI. Проведение кластеризации эколого-геохимических наблюдений на основе факторного анализа содержаний химических элементов.

С использованием пакета **STATISTICA 6.1** данная работа может быть выполнена в следующей последовательности.

На предварительном этапе создается (импортируется из Excel или набирается в Statistica) файл (с названием, напр., ИНАА-ISP почва.sta) с категориальной и отдельной по методам базой данных (см. Приложение П₁ и рис.2.1). При этом названия микроэлементов выносятся в названия переменных (столбцов).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Na _и	Fe _и	Ba _и	Sc _и	Cr _и	Co _и	La _и	Na ₁	Fe ₁	Ba ₁	Sc ₁	Cr ₁	Co ₁	La ₁	МЕТОД	Na
1	1,9	3,2	350	13,2	157,5	15,9	32,4	2,16	3,16	370	10,2	144,5	21,1	28	ИНАА	1,9
2	2	3,5	530	13,8	118,6	16,7	34,2	2,27	3,22	350	9,1	161,2	21,6	26,2	ИНАА	2
3	1,9	3,7	460	14,5	106,2	16,1	32,4	2,32	3,61	330	10	127,7	23	24,9	ИНАА	1,9
4	1,9	3,9	480	15,3	137,6	17,5	35,3	2,54	3,82	380	13,4	136,7	22,6	33,1	ИНАА	1,9
5	2,1	3,7	370	14,4	111,6	16,3	34,9	2,46	3,26	320	9,7	136,4	22,1	27,5	ИНАА	2,1
6	1,8	3	450	12,4	184,2	13,5	30,8	1,99	2,78	420	10,2	213,5	16,8	26,2	ИНАА	1,8

Рис.2.1. Фрагмент числового файла в рабочем окне STATISTICA

2.1. Описательная статистика

2.1.1. Вычисление оценок числовых характеристик содержаний химических элементов

На этапе I для вычисления оценок числовых характеристик содержаний химических элементов в ассоциациях по двум методам в активизированном пакете Statistica 6.1 необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль .

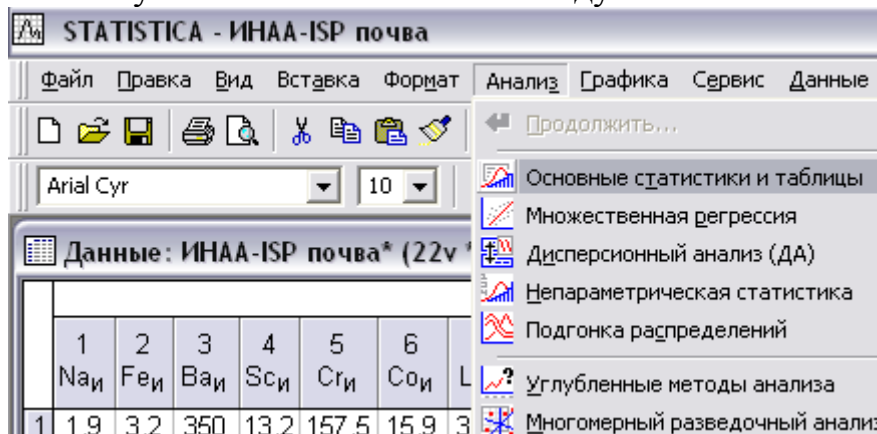
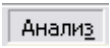
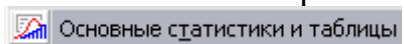


Рис.2.2. Запуск модуля  в системе STATISTICA

Шаг 2. В стартовой панели модуля нажмите клавишу .



Шаг 3. В появившемся окне выберите .

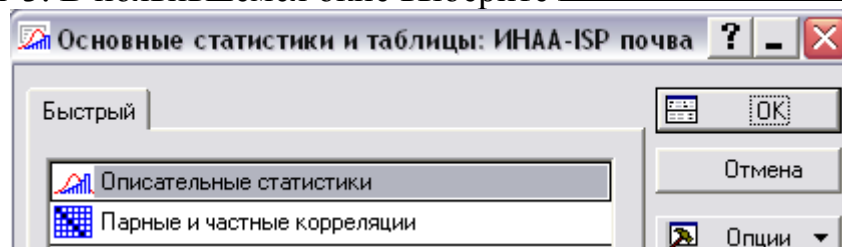
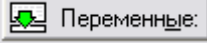
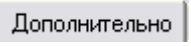


Рис.2.3. Окно  в STATISTICA.

Шаг 4. В открывшемся окне нажмите кнопку , выберите переменные раздельной по методам базы данных и нажмите ОК.

Шаг 5. Нажмите кнопку  и закажите (поставьте галочки) все нужные числовые характеристики (рис.2.4): объем выборки N , стандартное отклонение S , минимум (**Min**) и максимум (**Max**), среднее (m), медиана (Me), мода (Mo), асимметрию (A), эксцесс (E) и их стандартные ошибки σ_A , σ_E , а также σ_m .

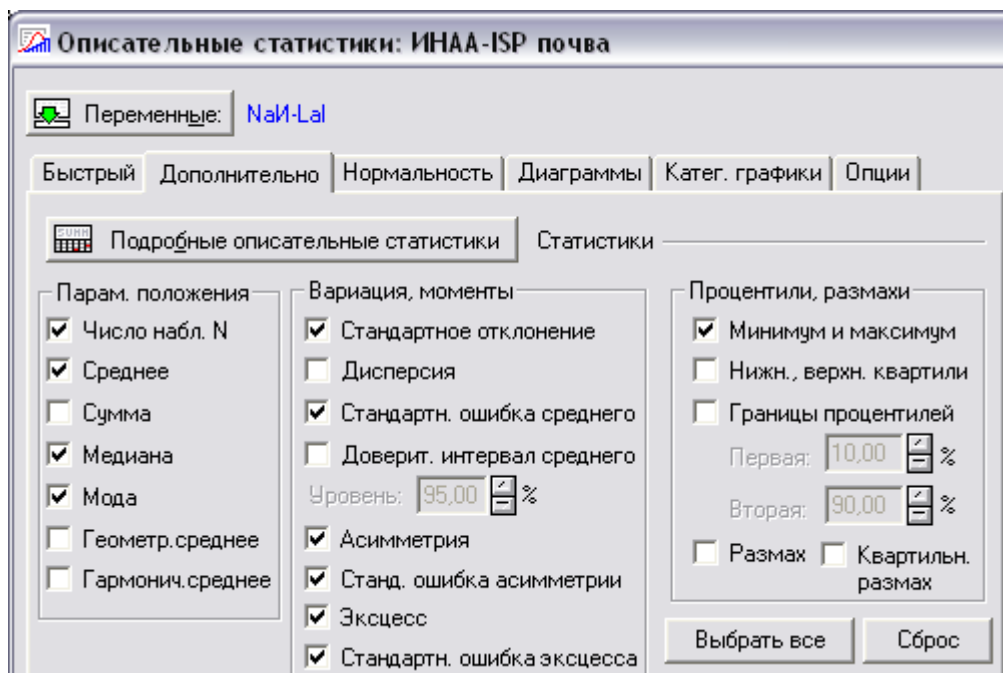


Рис.2.4. Режим **Дополнительно** в окне **Описательные статистики**.

Шаг 6. Нажмите в окне **Описательные статистики** кнопку



Следующая таблица появится на экране (табл.2.1):

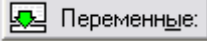
Таблица 2.1


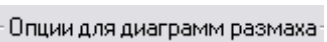
Числовые характеристики содержаний микроэлементов в почве населенных пунктов Томского региона, полученных методами ИНАА и ISP.

	<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
Na_И	91	1,33	1,30	1,30	0,50	2,30	0,42	0,04	0,31	0,25	-0,87	0,50
Fe_И	91	3,20	3,30	3,20	1,50	4,80	0,68	0,07	-0,77	0,25	0,42	0,50
Ba_И	91	501,1	480,0	400,0	260,0	970,0	137,4	14,40	0,88	0,25	0,81	0,50
Sc_И	91	12,08	12,80	12,80	5,40	16,80	2,88	0,30	-0,76	0,25	-0,28	0,50
Cr_И	91	127,0	117,4	Много	46,50	354,2	51,15	5,36	2,13	0,25	6,26	0,50
Co_И	91	15,79	15,30	14,30	3,70	64,40	6,99	0,73	4,23	0,25	27,21	0,50
La_И	91	27,03	27,50	Много	14,00	41,10	5,04	0,53	-0,21	0,25	0,08	0,50
Na_Г	91	1,79	1,697	1,99	0,88	2,96	0,49	0,05	0,65	0,25	-0,19	0,50
Fe_Г	91	3,04	3,048	Много	0,99	4,57	0,67	0,07	-0,57	0,25	1,73	0,50
Ba_Г	91	406,6	387,0	380,0	245,0	883,0	92,36	9,68	2,48	0,25	9,55	0,50
Sc_Г	91	9,18	9,00	Много	3,00	14,30	2,43	0,26	0,10	0,25	0,08	0,50
Cr_Г	91	117,5	104,0	80,0	62,90	360,0	49,50	5,19	2,76	0,25	10,50	0,50
Co_Г	91	18,65	19,00	Много	3,00	65,00	8,84	0,93	2,67	0,25	13,04	0,50
La_Г	91	22,21	24,00	20,0	2,00	33,10	6,92	0,73	-0,71	0,25	0,27	0,50

2.1.2. Построение диаграммы размаха

Для построения диаграммы размаха (графика ящички-усы), геометрически интерпретирующей набор числовых характеристик (центр и диапазон значений выбранной переменной), необходимо выполнить последовательность первых трех шагов п. 2.1.1.

Шаг 4. В открывшемся окне нажмите кнопку  и выберите переменную, например, N_{aH} , и нажмите ОК.

Шаг 5. Нажмите кнопку  и закажите тип графика выбором , например, Медиана/Квартиль/Размах (Медиана/квартильный размах(25%,75% процентиля)/размах (минимум, максимум))

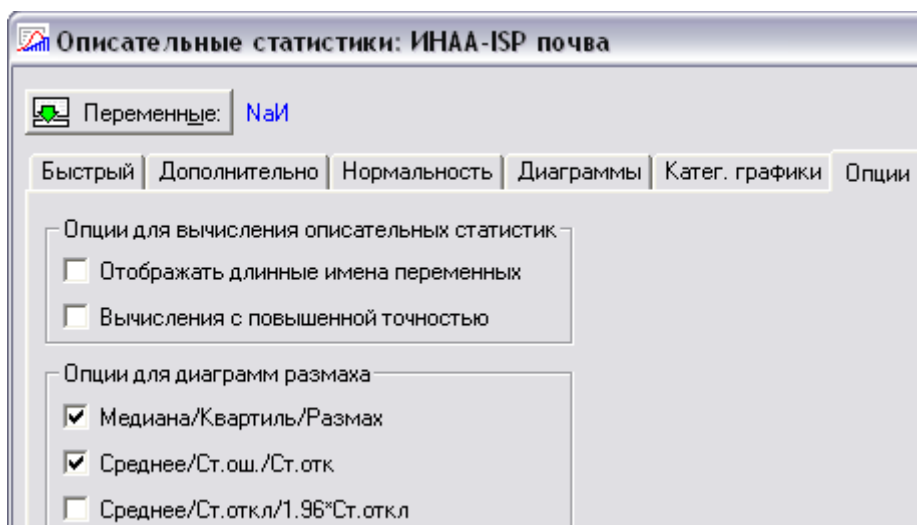
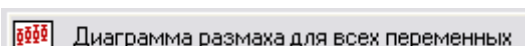
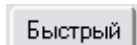


Рис.2.5.Режим  в окне .

Шаг 6. Закажите  в режиме 

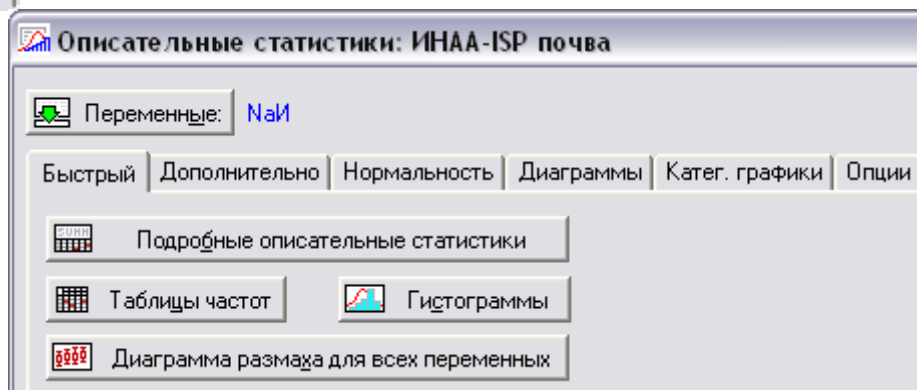


Рис.2.6.Режим  в окне .

Следующая картинка появится на экране (рис.2.7)

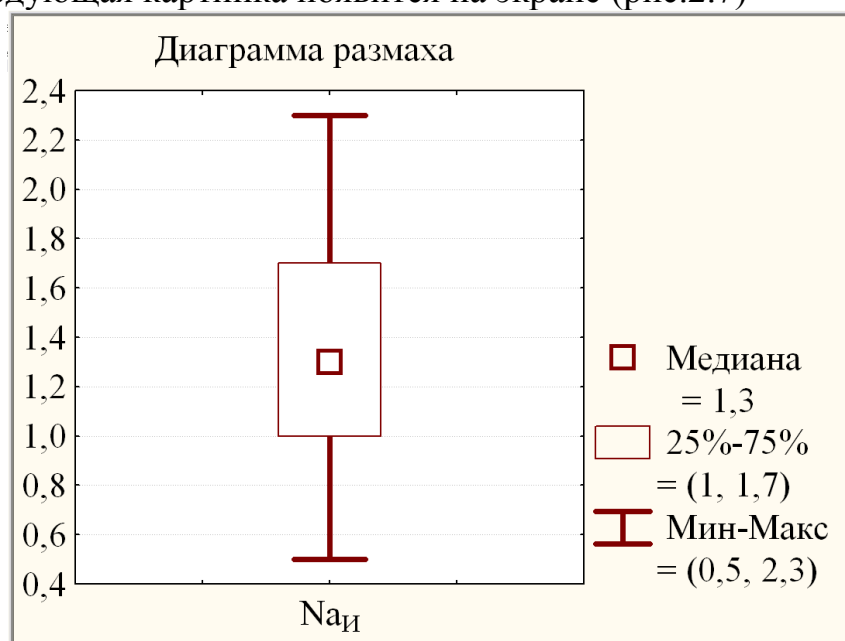


Рис.2.7. Диаграмма размаха микроэлемента NaI типа Медиана/ ...

Если на пятом шаге заказать тип графика выбором , например, Среднее/Ст.ош./Ст.отк (Среднее/ стандартная ошибка/ стандартное отклонение), то на экране появится следующая картинка

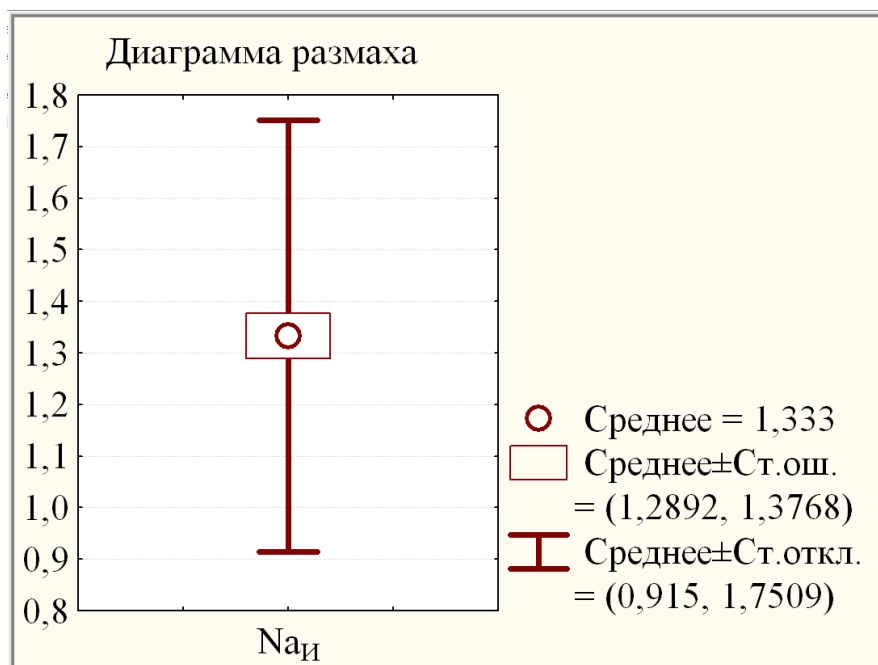



Рис.2.8. Диаграмма размаха микроэлемента NaI типа Среднее/ ...

На графике может быть представлено более одной переменной, если на четвертом шаге в окне  выбрать несколько переменных, например, Co_{II} и La_{II} .

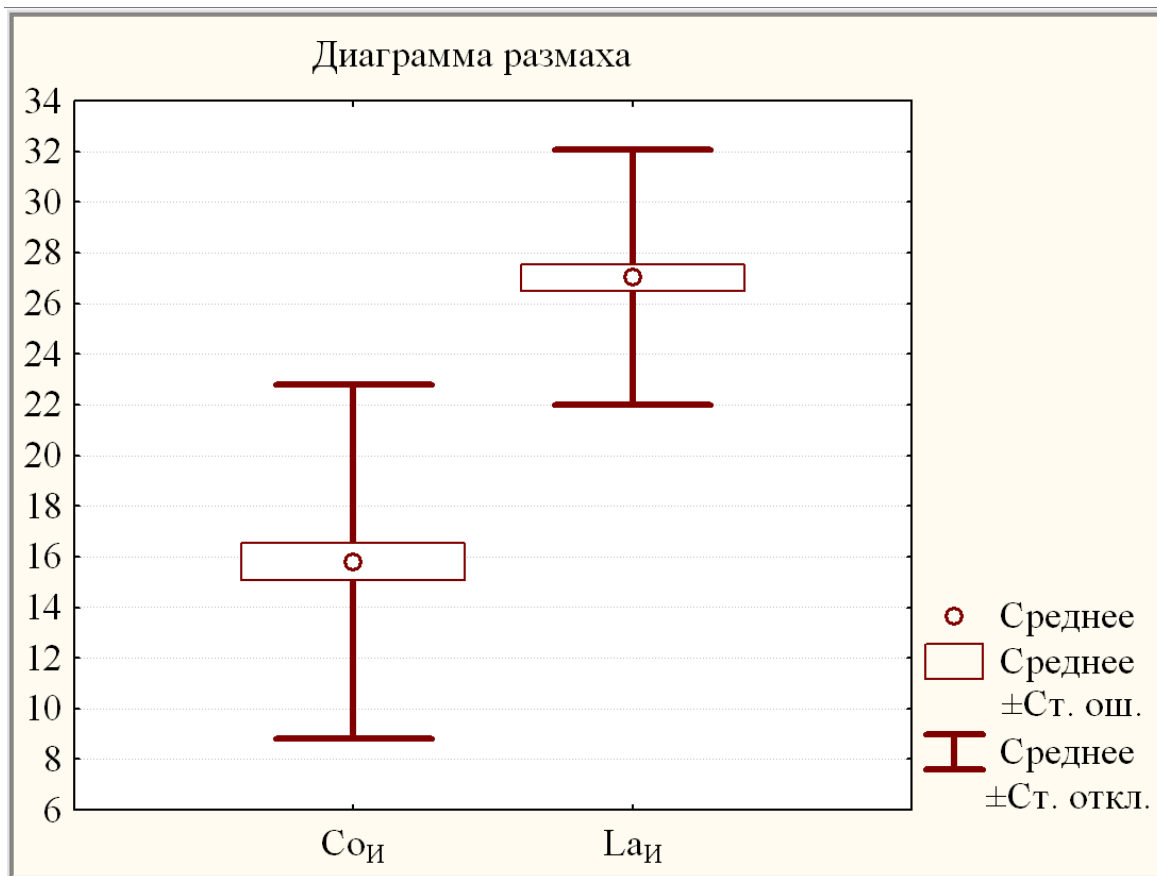


Рис.2.9.Составная диаграмма размаха Co_{II} и La_{II} типа Среднее/ ...

2.1.3. Построение гистограммы содержаний микроэлементов

Гистограмма (столбчатый график), геометрически отображающая распределение частот значений переменной по интервалам (см. п. 1.2.1; рис.1.16), дает наглядное представление о форме распределения и числовых характеристиках содержаний одной (простая гистограмма) или нескольких микроэлементов (составная гистограмма). Гистограмма наглядно показывает, какие значения или диапазон значений исследуемой переменной являются наиболее частыми, насколько сильно они различаются между собой, как сконцентрировано большинство наблюдений вокруг среднего, является распределение симметричным или нет, имеет ли оно одну моду или несколько мод, то есть является мультимодальным.

В системе STATISTICA существует несколько способов построения гистограмм. Например

Шаг 1. Запустите в головном меню (полоса меню) модуль .

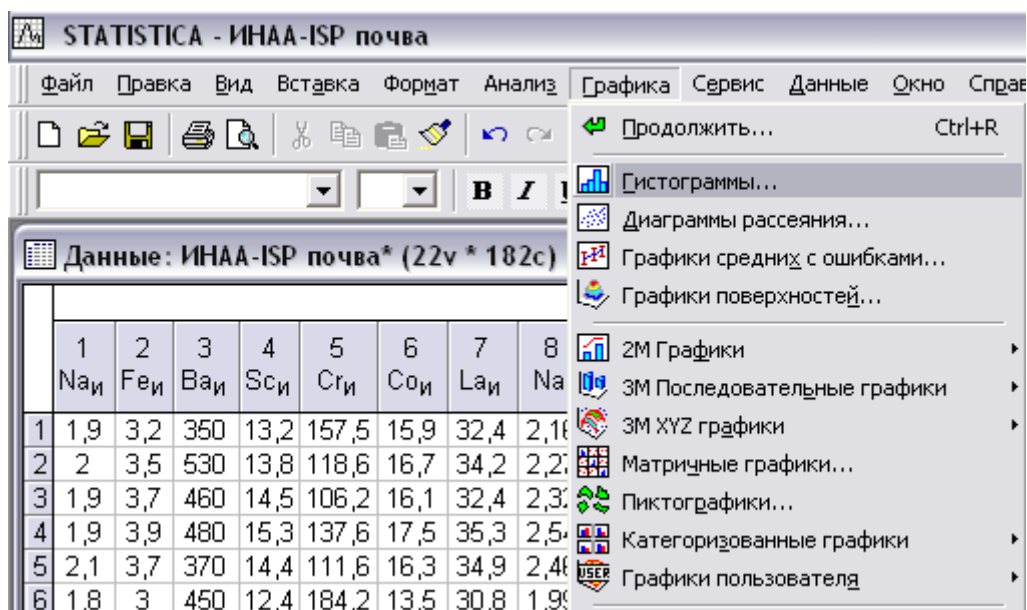


Рис.2.10. Запуск в головном меню модуля  в STATISTICA

Шаг 2. В стартовой панели модуля выберите .

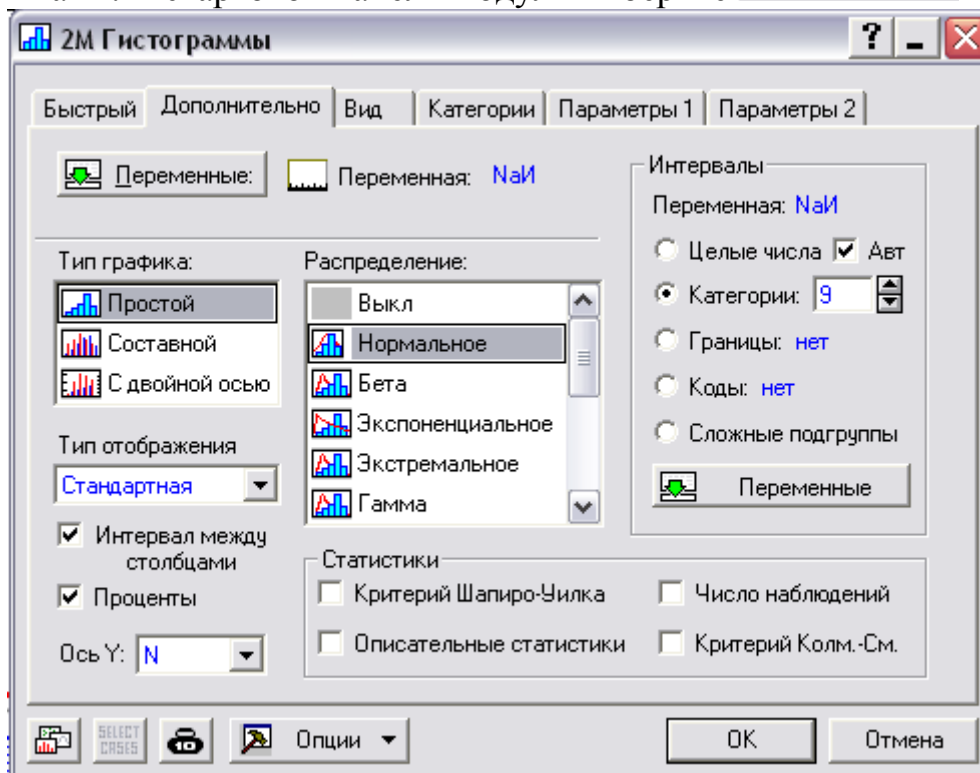
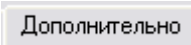



Рис.2.11. Режим  в окне 

Шаг 3. В режиме Дополнительно выберите параметры, как это показано на рис.2.11. Следующая картинка появится на экране.

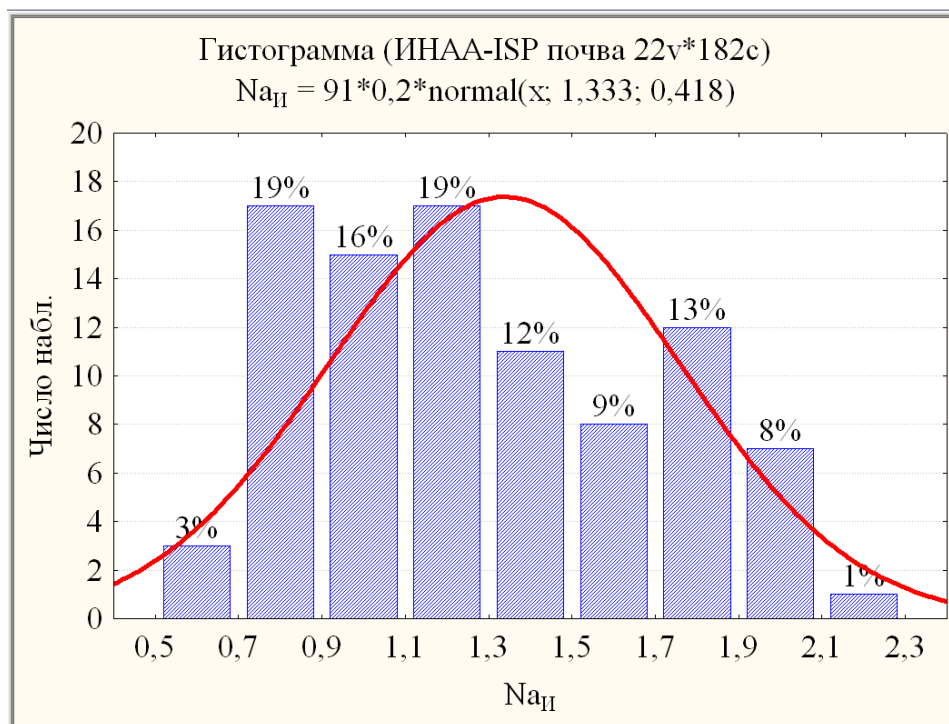


Рис.2.12. Гистограмма содержаний микроэлемента NaII

Если в окне Тип графика: выбрать Составной и заказать переменные NaII и NaI (рис.2.13),

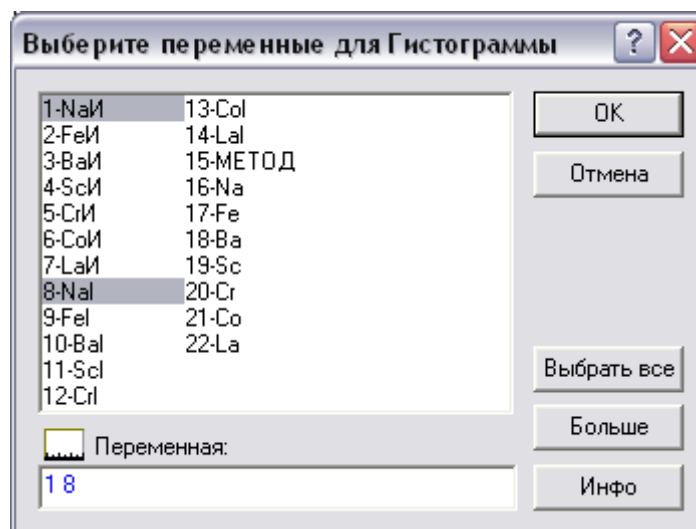


Рис.2.13. Окно Переменные: для типа графика Составной.

то на экране появится составная гистограмма выбранных переменных

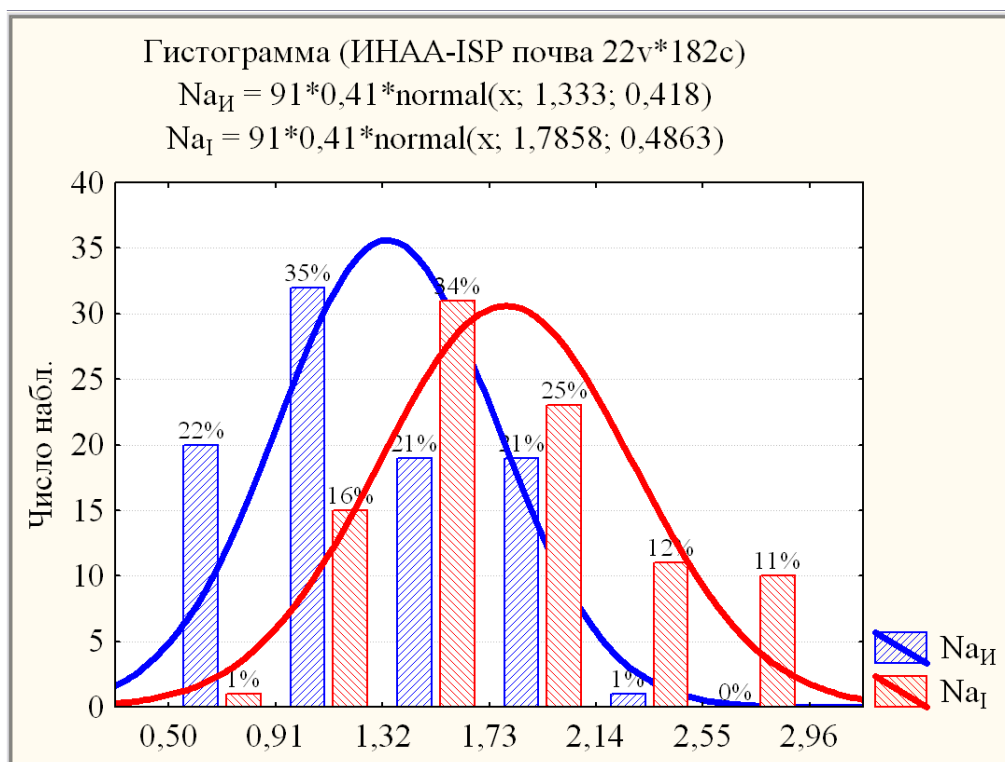


Рис.2.14. Составная гистограмма содержаний Na_{II} и Na_{I}

Для построения составных гистограмм можно использовать категориальную базу данных (см. рис. 2.15):

	15 МЕТОД	16 Na	17 Fe	18 Ba	19 Sc	20 Cr	21 Co	22 La
89	ИНАА	2,3	3,1	390	12,8	88,3	13,9	31,7
90	ИНАА	1,7	3,4	370	14,2	84,1	15,1	33,4
91	ИНАА	1,5	3,3	400	15,5	175	16,6	27,5
92	ISP	2,16	3,16	370	10,2	144,5	21,1	28
93	ISP	2,27	3,22	350	9,1	161,2	21,6	26,2
94	ISP	2,32	3,61	330	10	127,7	23	24,9

Рис.2.15. Фрагмент категориальной базы данных ИНАА–ISP почва

В связи с этим необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль **Графика** (рис.2.10).

Шаг 2. Выберите **Категоризованные графики** в стартовой панели модуля и далее щелкните мышкой на разновидности **Гистограммы...**

Шаг 3. В появившемся окне **2М Категоризованные гистограммы** щелкните **Дополнительно** и выберите параметры, как показано на рис.2.16:

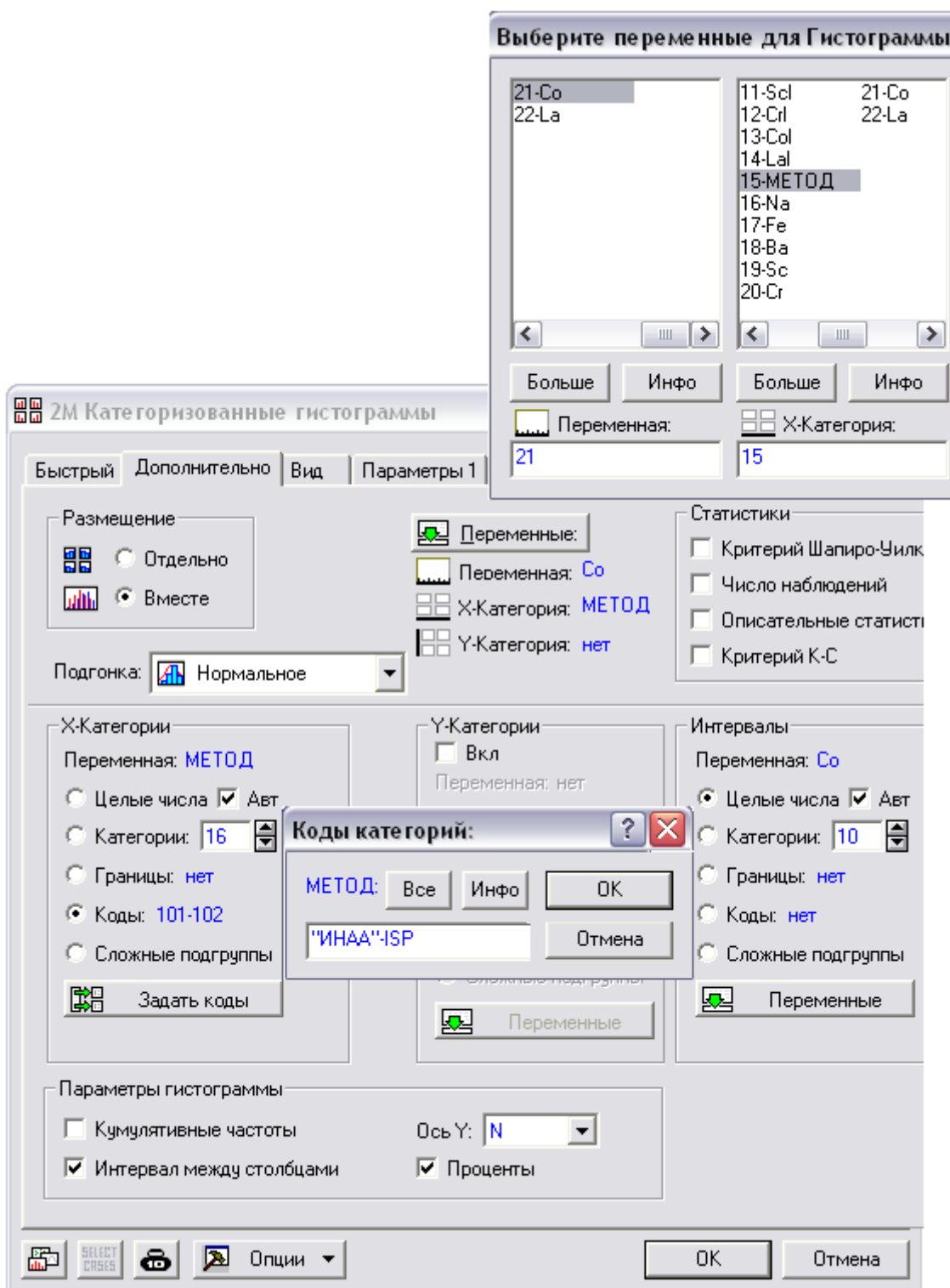


Рис. 2.16. Окно **2М Категоризованные гистограммы**

Шаг 4. Нажмите кнопку и выберите переменные категориальной базы данных (например, переменная «Со» – в окне и переменная «Метод» – в).

Шаг 5. Далее в окне выделите режим , нажмите кнопку , выберите в и нажмите ОК.

Шаг 6. После этого в окне выделите режим , в окне выделите режимы и , в окне выделите режим , в окне выберите (распределение) и нажмите ОК.

Составная гистограмма распределения содержания Co по методам ИНАА и ISP появится на экране (рис.2.17). Дополнительно построены соответствующие кривые распределения по нормальному закону.

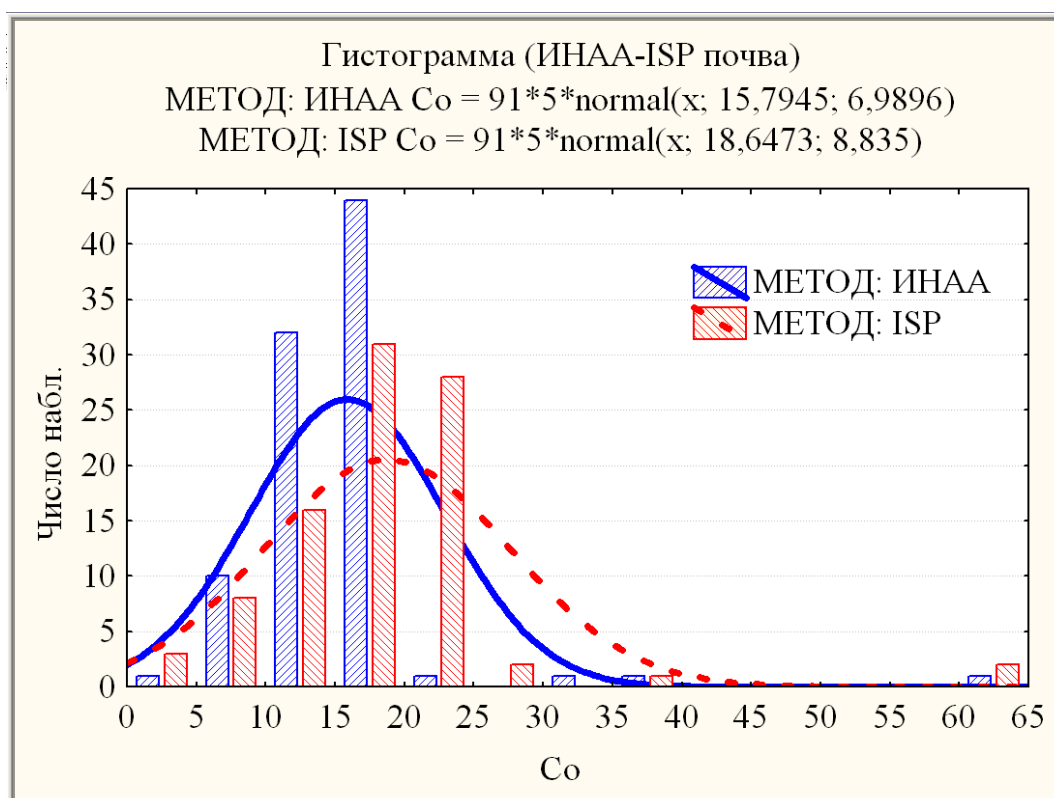


Рис. 2.17. Составная гистограмма распределения содержания Co по методам ИНАА и ISP

2.2. Проверка статистических гипотез

2.2.1. Проверка гипотезы о законе распределения содержаний химических элементов

В первой части **этапа II** для корректного выбора критерия сравнения двух выборок необходимо предварительно проверить гипотезу о нормальном распределении содержаний химических элементов. Существует несколько критериев и способов их реализации в системе STATISTICA. Например, при построении гистограммы (см. рис. 2.12) можно заказать проверку гипотезы о нормальном распределении содержания La_{II} , если в окне «Статистики» (см. рис. 2.11) пометить галочками тесты Критерий Шапиро-Уилка и Критерий Колм.-См. При этом реализуется также тест Лиллиефорса. Следующая картинка (рис.2.18) появится на экране. Согласно рис.2.18, распределение содержания химического элемента La_{II} незначимо (n.s.) отличается от нормального закона ($p \approx 0,72 > 0,10$ по критерию Шапиро – Уилка)

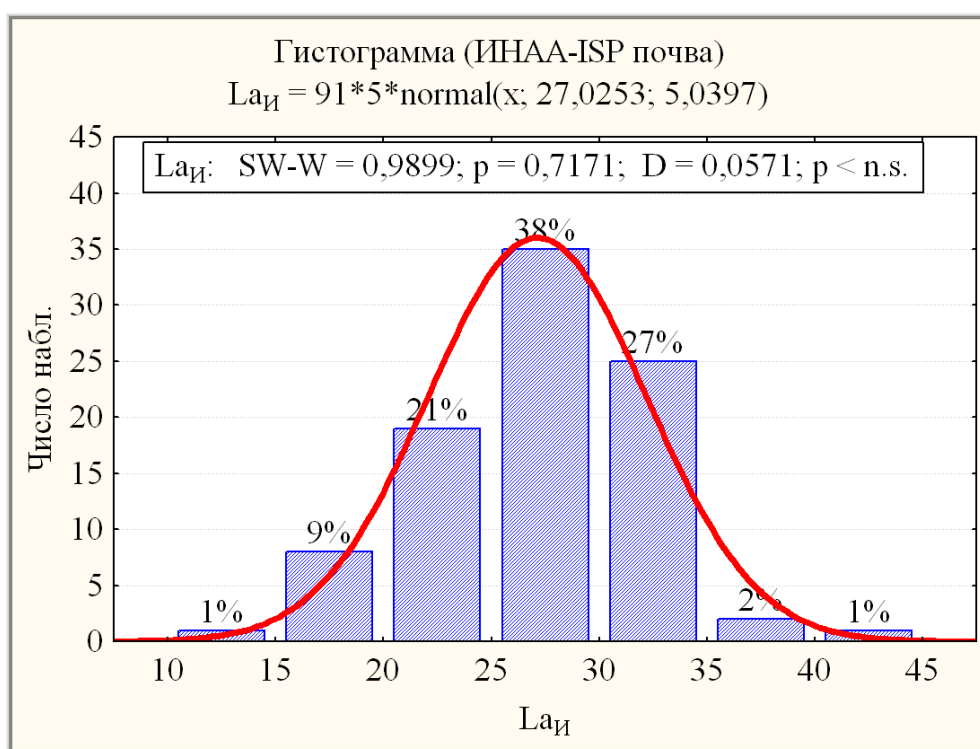






Рис.2.18. Проверка гипотезы о нормальном распределении La_{II} по критериям Шапиро – Уилка и Колмогорова – Смирнова

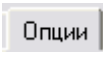
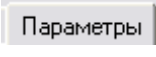
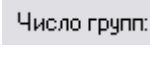
Для использования критерия хи – квадрат Пирсона необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль **Анализ** (см.рис.2.2).

Шаг 2. В стартовой панели модуля нажмите .

Шаг 3. В появившемся окне выберите  (распределение) и нажмите .

Шаг 4. В открывшемся окне нажмите кнопку  и выберите переменную L_{aI} и нажмите .

Шаг 5. Нажмите кнопку  (рис.2.19) и закажите (поставьте галочкой) критерий Хи – квадрат (здесь же можно заказать (поставить точкой) критерий Колмогорова – Смирнова), также нажмите кнопку  (рис.2.19), и установите в окне  число k столбцов гистограммы ($k=1+4lgn$, что дает $k \approx 9$ для объема выборки $n = 91$).

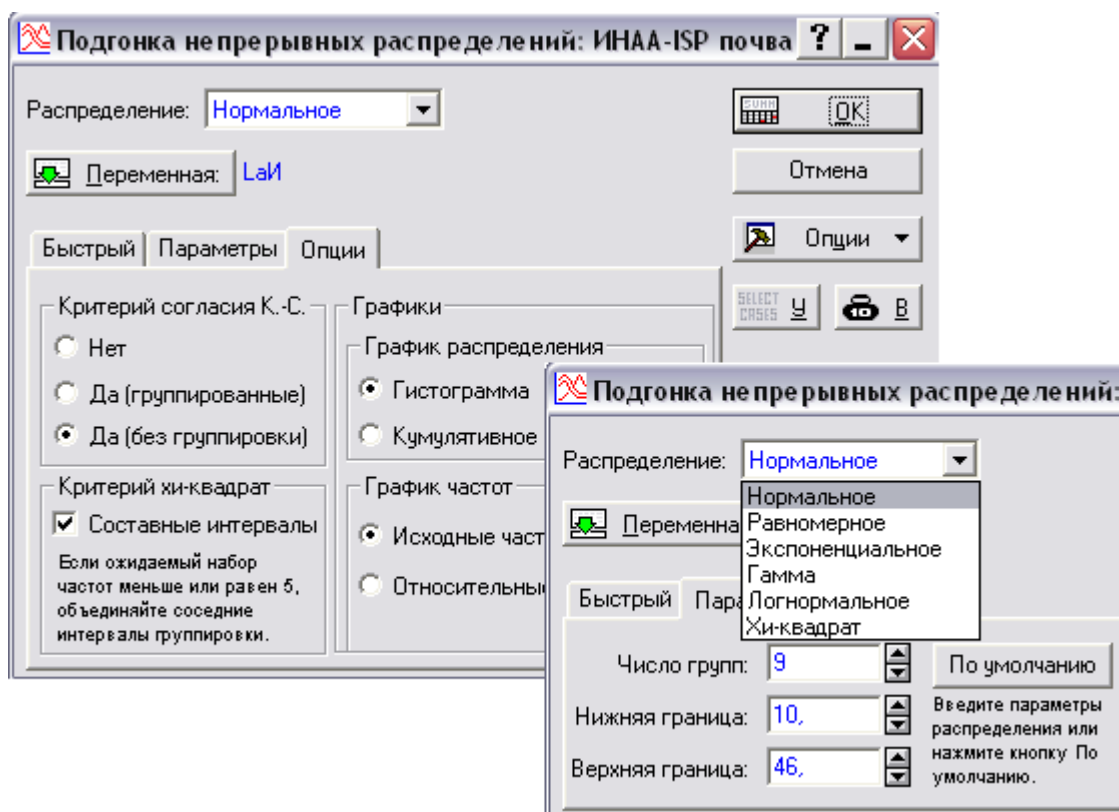

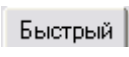
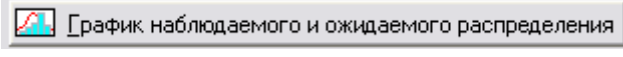


Рис.2.19. Выбор критерия Хи – квадрат и параметров гистограммы в окне 

Шаг 6. После чего перейдите в режим  и нажмите кнопку . Следующая картинка (рис.2.20) появится на экране:

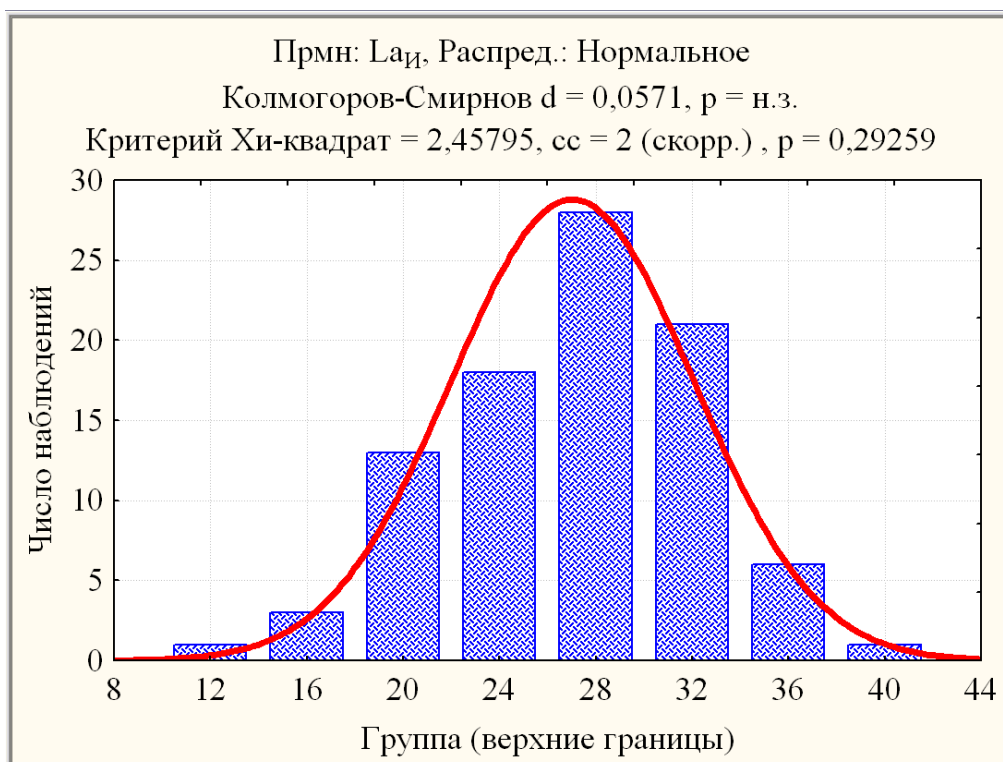


Рис. 2.20. Гистограмма распределения содержания La_{II} с кривой плотности нормального распределения и результатами теста Хи-квадрат

В подзаголовке указаны уровни значимости p критериев Хи-квадрат ($p \approx 0,29 > 0,10$) и Колмогорова – Смирнова ($p = \text{н.з.}$) незначимого отличия гистограммы от кривой плотности нормального распределения. Степень значимости отличия 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 нормальному теоретическому используют отношения показателей асимметрии \tilde{A} и эксцесса \tilde{E} к их стандартным ошибкам $\sigma_{\tilde{A}}$ и $\sigma_{\tilde{E}}$ соответственно ($\tilde{t}_1 = \frac{\tilde{A}}{\sigma_{\tilde{A}}}$, $\tilde{t}_2 = \frac{\tilde{E}}{\sigma_{\tilde{E}}}$). Так как эти отношения по абсолютной величине не превышают 3 (см. табл.2.1), то нет оснований отвергать гипотезу о нормальном распределении La_{II} .

Ниже приведен еще один пример: гистограмма распределения содержания $Сг_{II}$ с кривыми плотностей нормального (рис.2.21) и логнормального (рис.2.22) распределений и результаты тестов Колмогорова-Смирнова и Хи-квадрат.

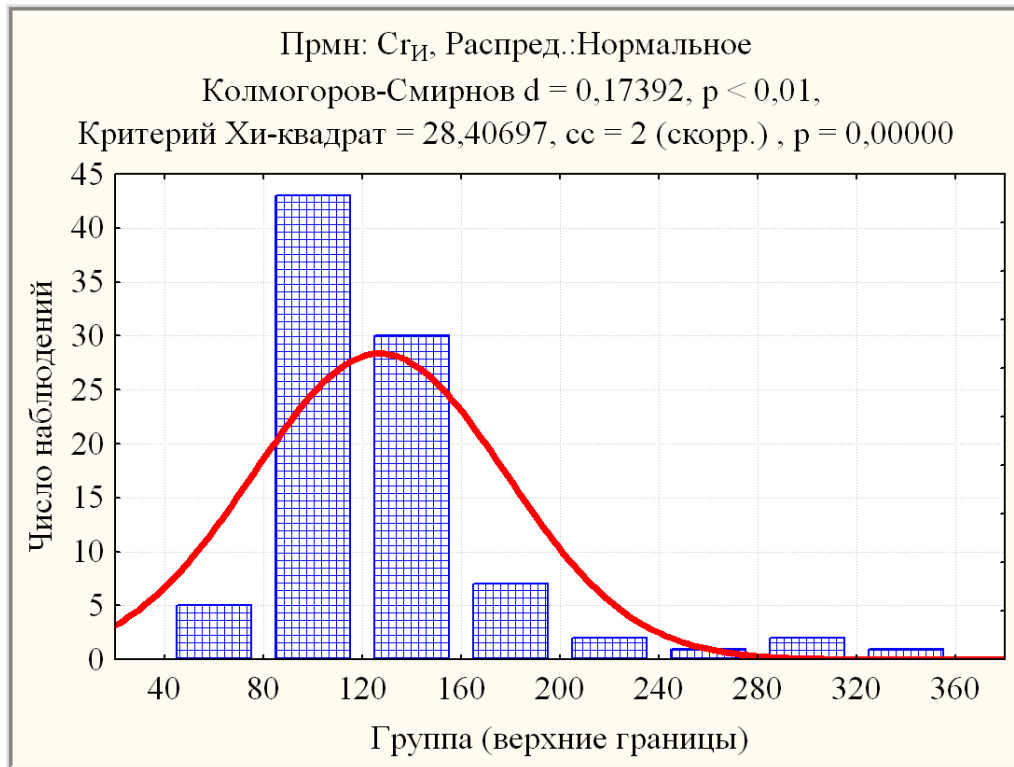


Рис. 2.21. Гистограмма распределения $C_{гИ}$ с кривой плотности нормального распределения и результатами тестов

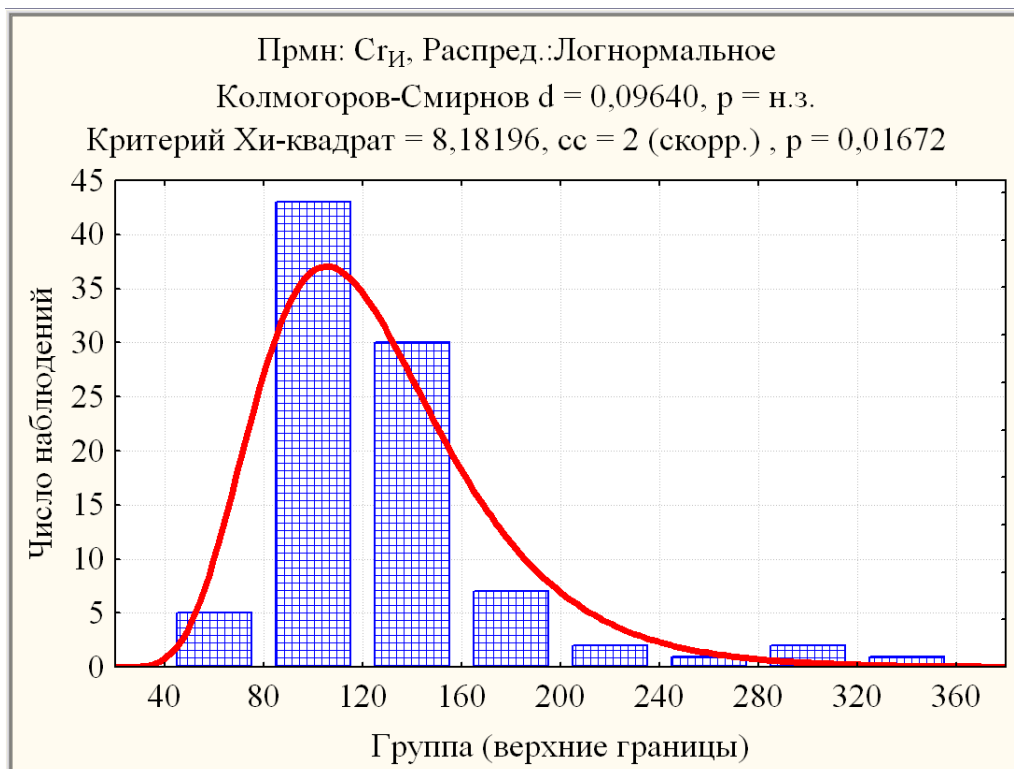


Рис. 2.22. Гистограмма распределения $C_{гИ}$ с кривой плотности логнормального распределения и результатами тестов

Так как распределение содержания $C_{гИ}$ статистически значимо отличается от нормального (рис.2.21), то в этом случае его можно сравнить с более подходящим теоретическим законом распределения, например, с логнормальным: см.рис. 2.19, где следует выбрать распределение Логнормальное. Результаты представлены на рис. 2.22

Критерий Колмогорова-Смирнова дает незначимое ($p = \text{н.з.}$) отличие распределения содержания $C_{гИ}$ от логнормального закона. Критерий Хи-квадрат Пирсона – статистически значимое ($0,050 > p \approx 0,017 > 0,005$). Последнее обстоятельство связано с тем, что при вычислении значения критерия Пирсона малочисленные интервалы (число наблюдений меньше пяти) правого хвоста, характерные для логнормального распределения, объединяются (см. замечание на рис.2.18) с восемью интервалов - столбцов (см. рис.2.22) до пяти, что приводит к числу степеней свободы, равному двум ($ss = 2$).

2.2.2. Оценка различия содержаний двух выборок

Во второй части этапа II для сравнения содержаний химических элементов, полученных при проведении эколого-геохимических исследований разными методами, следует применять в STATISTICA критерий Стьюдента в случае нормально распределенных содержаний химических элементов и непараметрические критерии Манна-Уитни и Колмогорова-Смирнова (Боровиков В.П., 2003) в остальных случаях.

Для начала сравнение выборок можно провести графически: аналогично рис.2.9 можно сравнить диаграммы размаха содержаний $La_{гИ}$ и $La_{гI}$ (рис. 2.23)

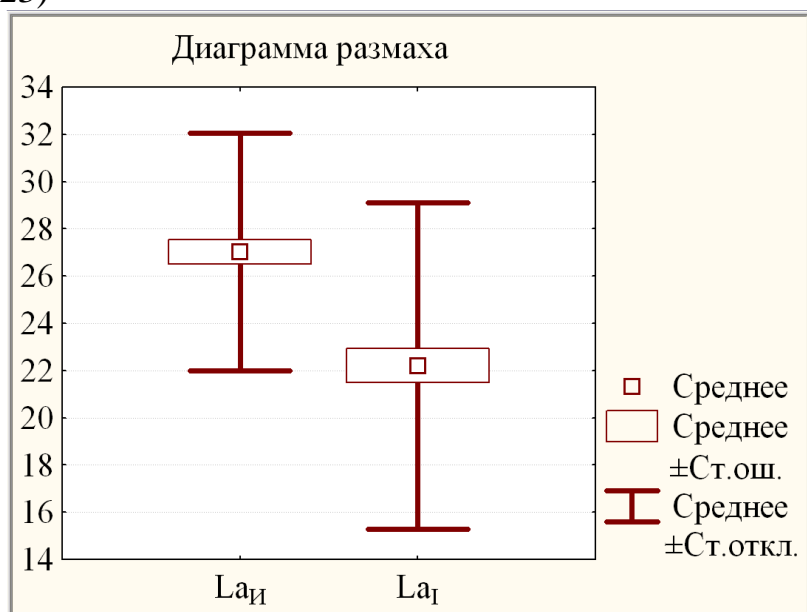


Рис. 2.23. Составная диаграмма размаха $La_{гИ}$ и $La_{гI}$

Для построения составной диаграммы размаха распределения содержаний La_{II} и La_I (рис.2.23) может быть использована категориальная база данных (рис.2.15) аналогично рис.2.16. Для этого в стартовой панели модуля Категоризованные графики вместо Гистограммы... щелкните на Диаграммы размаха... для входа в 2М Категоризованные диаграммы размаха, где с помощью кнопки Переменные: выберите переменные категориальной базы данных (например, переменная «La» – в окне Зависимая переменная: и переменная «Метод» – в X-Категория:) а в окне Подгонка: выберите Выкл и нажмите ОК.

Аналогично рис.2.14 сравнение выборок La_{II} и La_I можно провести графически по составным гистограммам (рис. 2.24). Попутно при построении гистограмм можно заказать проверку гипотезы о нормальном распределении содержания микроэлементов тестами Шапиро – Уилка и Колмогорова – Смирнова (рис.2.24).

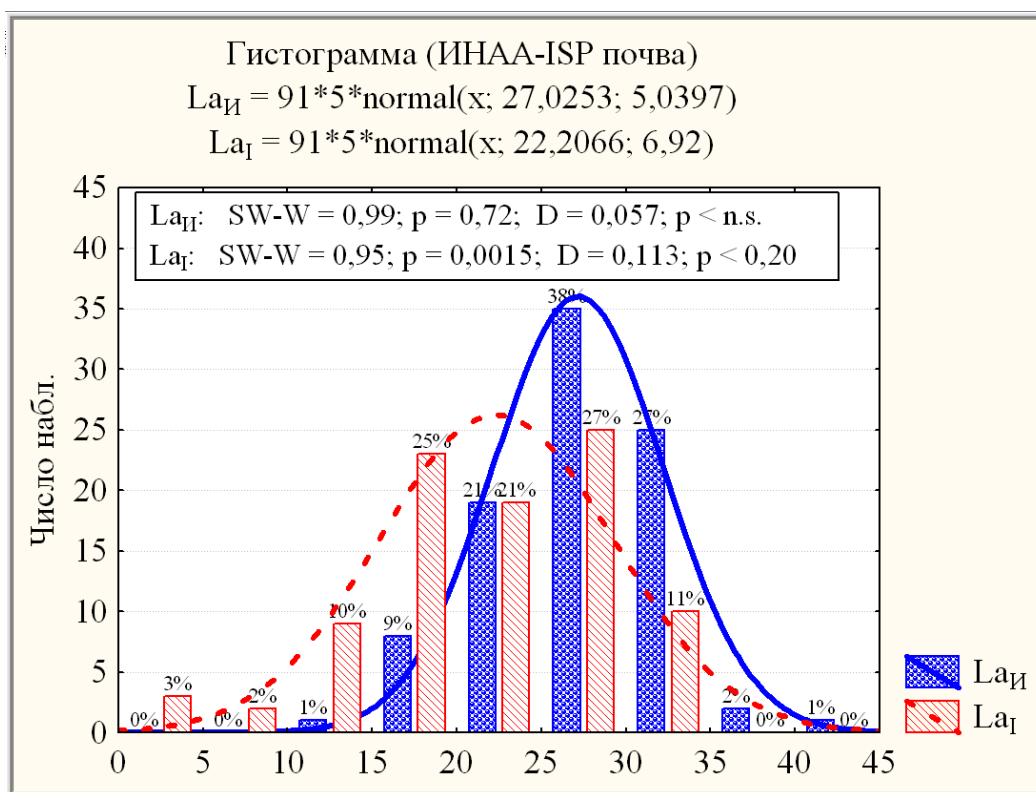


Рис. 2.24. Составная гистограмма распределения La_{II} и La_I

Для проведения аналитической оценки различия содержаний химических элементов La_{II} и La_I по критерию Стьюдента необходимо выполнить последовательность следующих действий:

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

Шаг 2. Нажмите **Основные статистики и таблицы** (рис.2.2-3).

Шаг 3. В появившемся окне выберите **t-критерий для независимых выборок** и нажмите **OK** (рис.2.25).

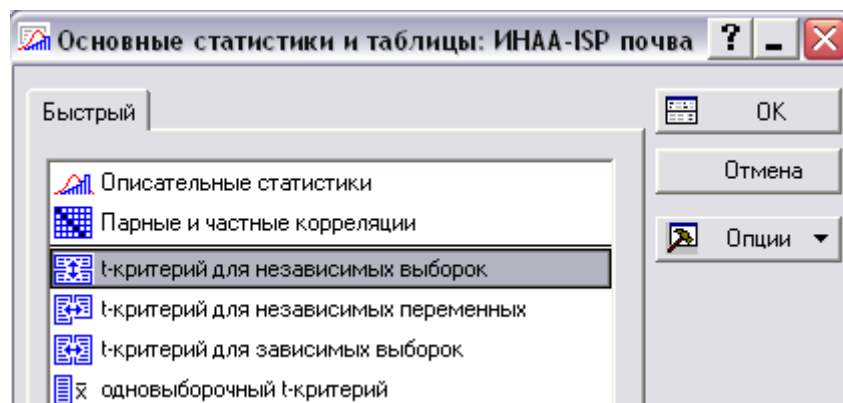


Рис.2.25. Выбор **t-критерий для независимых выборок**

Шаг 4. В открывшемся окне нажмите кнопку **Переменные:** и выберите переменные категориальной базы («Метод» - группирующая, а химические элементы - зависимые) и нажмите **OK**. После этого в окнах кодов для групп появятся «ИНАА» и «ISP»

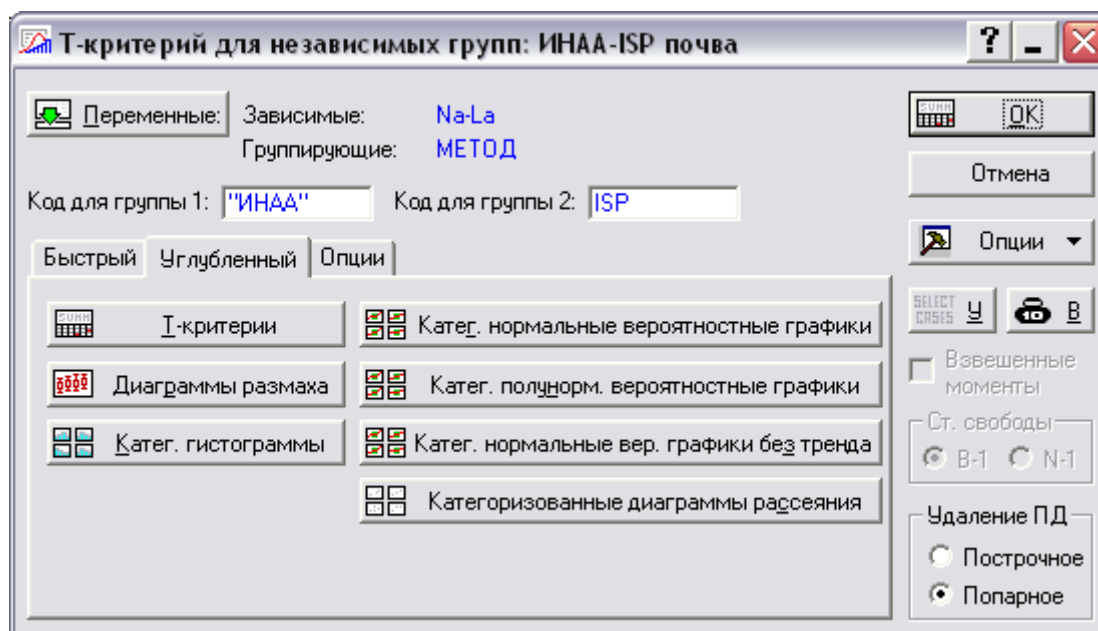


Рис.2.26. Окно **Т-критерий для независимых групп:**

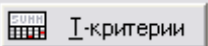
Шаг 5. После этого нажмите кнопку . Следующая таблица с результатами появится на экране:


Таблица 2.2


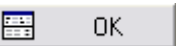
Параметрические t - критерий Стьюдента (и F - критерий Фишера) сравнения содержаний микроэлементов, полученных разными методами ИНАА и ISP.

	m_{II}	m_I	t -знач.	сс	p_t	N_{II}	N_I	S_{II}	S_I	F -знач.	p_F
Na	1,33	1,79	-6,74	180	0,000	91	91	0,42	0,49	1,35	0,153
Fe	3,20	3,04	1,55	180	0,124	91	91	0,68	0,67	1,03	0,904
Ba	501,10	406,63	5,44	180	0,000	91	91	137,39	92,36	2,21	0,000
Sc	12,08	9,18	7,33	180	0,000	91	91	2,88	2,43	1,40	0,112
Cr	127,01	117,51	1,27	180	0,205	91	91	51,15	49,50	1,07	0,757
Co	15,79	18,65	-2,42	180	0,017	91	91	6,99	8,84	1,60	0,027
La	27,03	22,21	5,37	180	0,000	91	91	5,04	6,92	1,89	0,003

Заметим, что корректное применение параметрического t - критерия Стьюдента предполагает выполнения двух условий: нормальность распределения сравниваемых выборок и не значимое различие дисперсий ($p_F > 0,10$ по F - критерию Фишера). В противном случае в системе STATISTICA реализуются непараметрические критерии:

Шаг 1. Запустите в головном меню модуль .

Шаг 2. Нажмите  **Непараметрическая статистика** в стартовой панели модуля (см. рис. 2.2).

Шаг 3. В появившемся окне выберите  **Сравнение двух независимых групп** (рис.2.27) и нажмите .

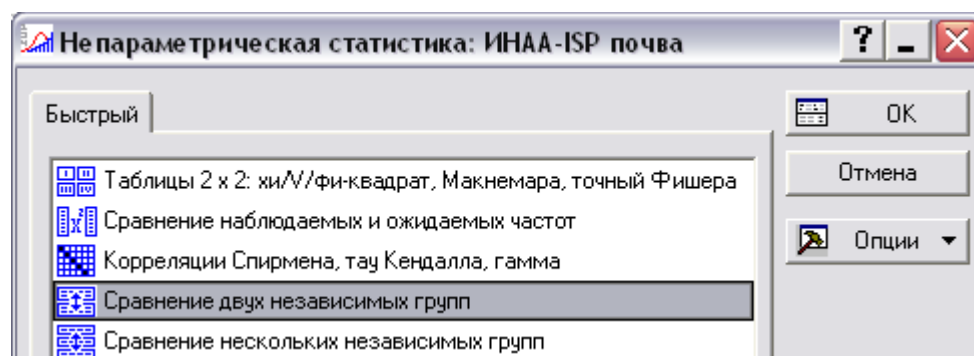




Рис.2.27. Выбор  **Сравнение двух независимых групп** в окне



Шаг 4. В открывшемся окне (рис.2.28) нажмите  Переменные и выберите переменные категориальной базы данных («Метод» - группировающая, а химические элементы - зависимые) и нажмите . После этого в окнах кодов для групп появятся «ИНАА» и «ISP»

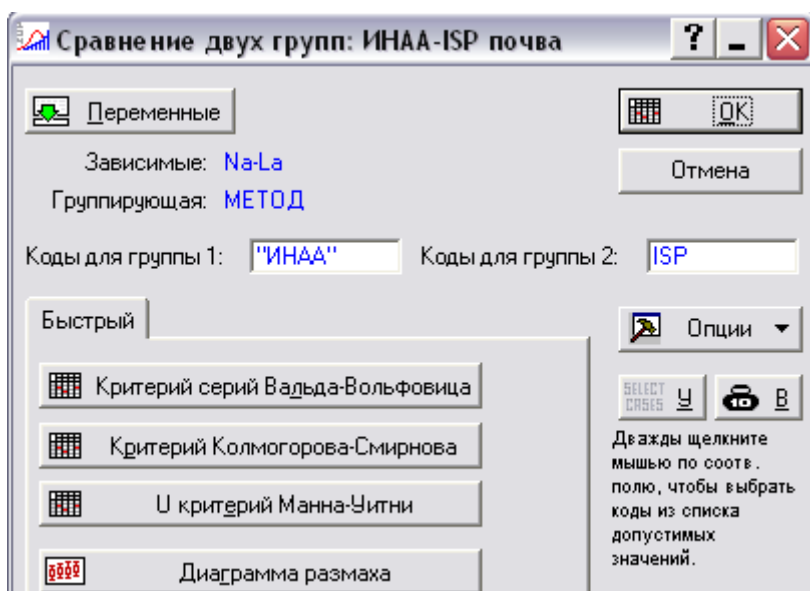


Рис.2.28. Окно  Сравнение двух групп


Шаг 5. После этого нажмите кнопку  Критерий Колмогорова-Смирнова. Следующая таблица с результатами появится на экране:

Таблица 2.3

Непараметрический критерий Колмогорова-Смирнова сравнения средних содержаний, полученных разными методами ИНАА и ISP

К-S	Макс.отр разн.	Макс.по разн.	p-уров.	m_{II}	m_I	S_{II}	S_I	N_{II}	N_I
Na	-0,451	0,000	$p < .001$	1,33	1,79	0,42	0,49	91	91
Fe	-0,066	0,264	$p < .005$	3,20	3,04	0,68	0,67	91	91
Ba	0,000	0,429	$p < .001$	501,1	406,6	137,39	92,36	91	91
Sc	0,000	0,549	$p < .001$	12,08	9,18	2,88	2,43	91	91
Cr	-0,044	0,176	$p > .10$	127,0	117,5	51,15	49,50	91	91
Co	-0,440	0,055	$p < .001$	15,79	18,65	6,99	8,84	91	91
La	0,000	0,319	$p < .001$	27,03	22,21	5,04	6,92	91	91

Основной интерес представляет уровень значимости (p -уров.) различий содержаний, полученных разными методами.

Шаг 6. Вернитесь в окно  и нажмите кнопку

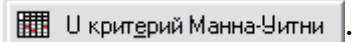
 :.

Таблица 2.4

Непараметрический U - критерий Манна - Уитни сравнения содержаний микроэлементов, полученных методами ИНАА и ISP

	Сум.ранг ИНАА	Сум.ранг ISP	U	Z	p -уров.	Z	p -уров.	N_{II}	N_I
Na	6272,0	10381,0	2086,0	-5,78	0,000	-5,78	0,000	91	91
Fe	9175,5	7477,5	3291,5	2,39	0,017	2,39	0,017	91	91
Ba	10257,5	6395,5	2209,5	5,43	0,000	5,44	0,000	91	91
Sc	10628,5	6024,5	1838,5	6,48	0,000	6,48	0,000	91	91
Cr	9052,5	7600,5	3414,5	2,04	0,041	2,04	0,041	91	91
Co	6826,0	9827,0	2640,0	-4,22	0,000	-4,22	0,000	91	91
La	9981,5	6671,5	2485,5	4,66	0,000	4,66	0,000	91	91

Здесь также основной интерес представляет столбец уровней значимости (p -уров.) различий содержаний микроэлементов, полученных методами ИНАА и ISP.

Согласно табл.2.2-4 содержания микроэлементов Na, Ba, Sc, и La, полученных разными методами ИНАА и ISP, различаются высоко значимо ($p < 0,001$). Графическое представление высоко значимых различий дают составные гистограммы Na_{II} и Na_I (рис.2.14), а также La_{II} и La_I (рис.2.24) и составная диаграмма размаха La_{II} и La_I (рис.2.23). Содержания Co, полученные разными методами, различаются также высоко значимо ($p < 0,001$), согласно корректным непараметрическим методам (нарушение нормальности распределения следует из рис.2.17 и табл.2.1 ($\tilde{t}_1, \tilde{t}_2 > 3$), а значимое различие дисперсий из табл.2.2

Результаты тестирования различий содержаний микроэлементов Fe_{II} и Fe_I не значимы по параметрическому критерию (табл.2.2) и значимы по непараметрическим критериям (табл.2.3 - 4). Но в связи с нарушением требования нормальности распределений содержаний Fe_{II} и Fe_I (см. табл.2.1 и рис.2.29) следует признать более корректными выводы по непараметрическим критериям.

Аналогичная ситуация по Cr (рис.2.30): не значимые различия содержаний микроэлементов Cr_{II} и Cr_I по параметрическому критерию (табл.2.2) и слабо значимые по непараметрическим критериям (табл.2.3 - 4).

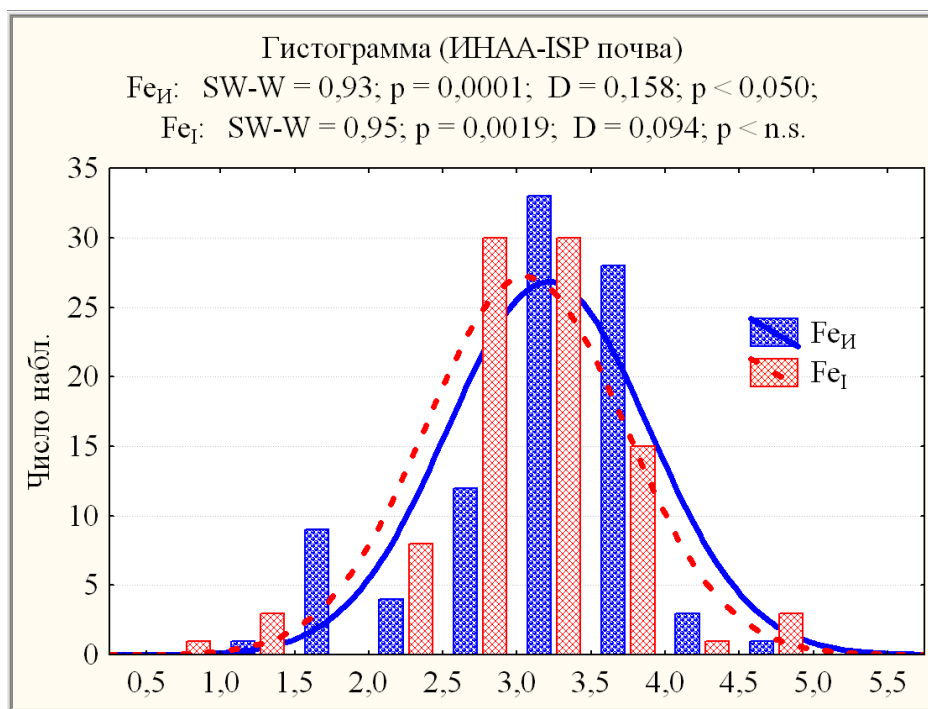


Рис. 2.29. Составная гистограмма распределения Fe_И и Fe_Г

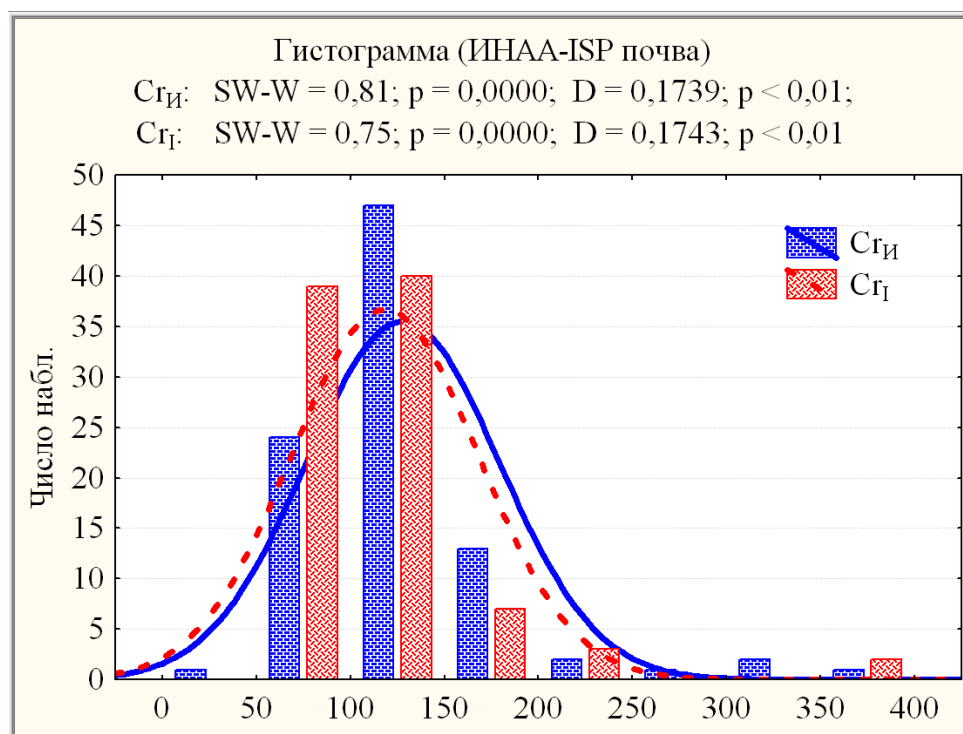





Рис. 2.30. Составная гистограмма распределения Cr_И и Cr_Г

В связи с более корректным непараметрическим подходом к оценке различий распределений содержаний микроэлементов Fe_И и Fe_Г, а также Cr_И и Cr_Г, полученных разными методами ИНАА и ISP, для графической иллюстрации этих различий можно использовать непарамет-

рические характеристики (медиана, квартильный размах(25%,75% процентиля) и абсолютный размах (минимум, максимум)) и построить с помощью кнопки  Диаграмма размаха (рис.2.28) на стадии применения  Не параметрическая статистика. В окне  Сравнение двух групп составные диаграммы размаха (рис.2.31–32).

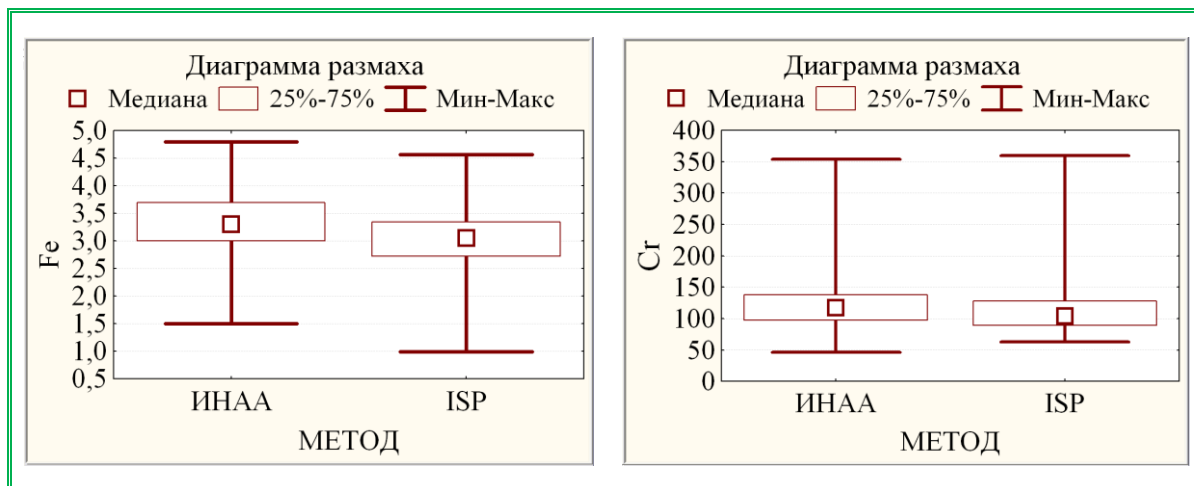


Рис. 2.31. Составные диаграммы размаха по методам ИНАА и ISP для Fe (слева) и Cr (справа)

По совокупности параметрического и непараметрический критериев примером сильно значимых различий распределений содержаний микроэлементов, полученных разными методами ИНАА и ISP, является Sc. Графически это иллюстрируется составными диаграммами размаха по параметрическим и непараметрическим характеристикам (рис.2.32), построенными аналогично рис.2.9.

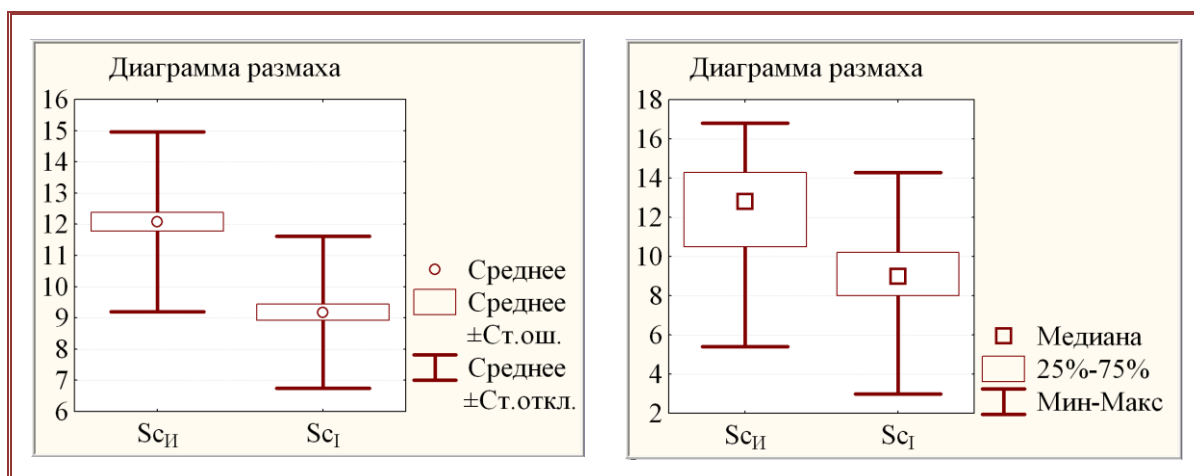


Рис. 2.32. Составные диаграммы размаха по методам для Sc по параметрическим (слева) и непараметрическим (справа) характеристикам

Сравнение выборок Sc_{II} и Sc_I можно провести также наглядно графически по составной гистограмме (рис.2.33), построенной аналогично рис.2.14.

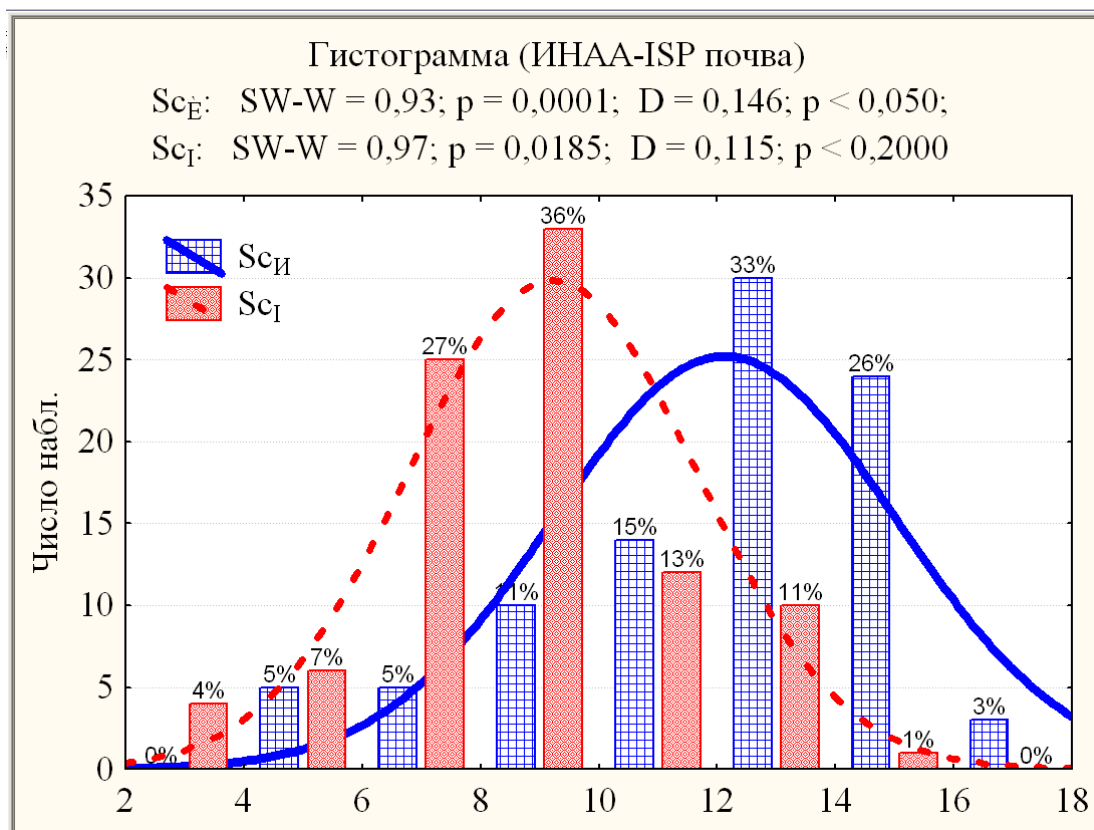




Рис. 2.33. Составная гистограмма распределения Sc_{II} и Sc_I

2.3. Компьютерный дисперсионный анализ

Как известно (п.1.2 3. Дисперсионный анализ), наиболее распространенной целью дисперсионного анализа является проверка статистической значимости различия между средними нескольких групп. Эта проверка проводится с помощью разбиения общей дисперсии (вариации) на части, одна из которых обусловлена случайной ошибкой $SS_{ошибки}$ (то есть внутригрупповой изменчивостью), а вторая $SS_{эффекта}$ связана с различием средних значений (то есть с межгрупповой изменчивостью). Проверка значимости в дисперсионном анализе основана на сравнении компоненты дисперсии, обусловленной межгрупповым разбросом (называемой средним квадратом эффекта $MS_{эффекта}$) и компоненты дисперсии, обусловленной внутригрупповым разбросом (называемой средним квадратом ошибки $MS_{ошибки}$). Если это различие значимо по F – критерию, нулевая гипотеза отвергается и принимается альтернативная гипотеза о существовании различия между средними.



В частном случае сравнения двух средних (п.2.2 2) можно было бы сразу вычислить t -критерий. Полученные результаты, естественно, совпадут с результатами дисперсионного анализа. Однако дисперсионный анализ содержит гораздо более гибкие и мощные технические средства, позволяющие проводить более сложные исследования. Например, дисперсионный анализ позволяет обнаружить эффекты взаимодействия между факторами и, поэтому, позволяет проверять более сложные гипотезы. Многофакторный дисперсионный анализ позволяет изучать каждый фактор, управляя значениями других факторов. Это, в действительности, и является основной причиной его большой статистической мощности (для получения значимых результатов требуется меньше объемы выборок). По этой причине дисперсионный анализ даже на небольших выборках дает статистически более значимые результаты, чем простой t -критерий.

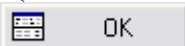
В пакете STATISTICA дисперсионный анализ выделен в отдельный модуль  Дисперсионный анализ (ДА), позволяющий решать различные типовые задачи дисперсионного анализа (п.1.2 3).

Рассмотрим применение модуля  Дисперсионный анализ (ДА) к примеру содержания химических элементов, полученных при проведении эколого-геохимических исследованиях разными методами (п.2.2 2) на основе базы данных (см. Приложения П₁), в которой кроме фактора МЕТОД (с двумя уровнями ИНАА и ISP) введен в рассмотрение второй фактор (СЕКТОР с тремя уровнями), характеризующий территориальную неоднородность эколого-геохимических измерений по трем уровням – секторам зон влияния Сибирского химического комбината: юго-западный (ЮЗ) сектор (измерения на территории с. Зоркальцево №1–19 в П₁ и с. Поросино №81–83), юго-восточный (ЮВ) сектор (измерения на территории г. Томска №20–80) и северо-восточный (СВ) сектор (измерения на территории сс. Георгиевка, Наумовка и Надежда №84–91).

2.3.1. Однофакторный дисперсионный анализ

Средствами однофакторного дисперсионного анализа проведем оценку значимости влияния фактора СЕКТОР на переменную содержания химических элементов (рис.2.34).

Шаг 1. Запустите  Анализ и щелкните  Дисперсионный анализ (ДА) (см. рис. 2.34).

Шаг 2. В появившемся окне выберите тип анализа (однофакторный дисперсионный анализ) и задайте метод (диалог быстрых спецификаций), как показано на рис.2.35, и нажмите  ОК.

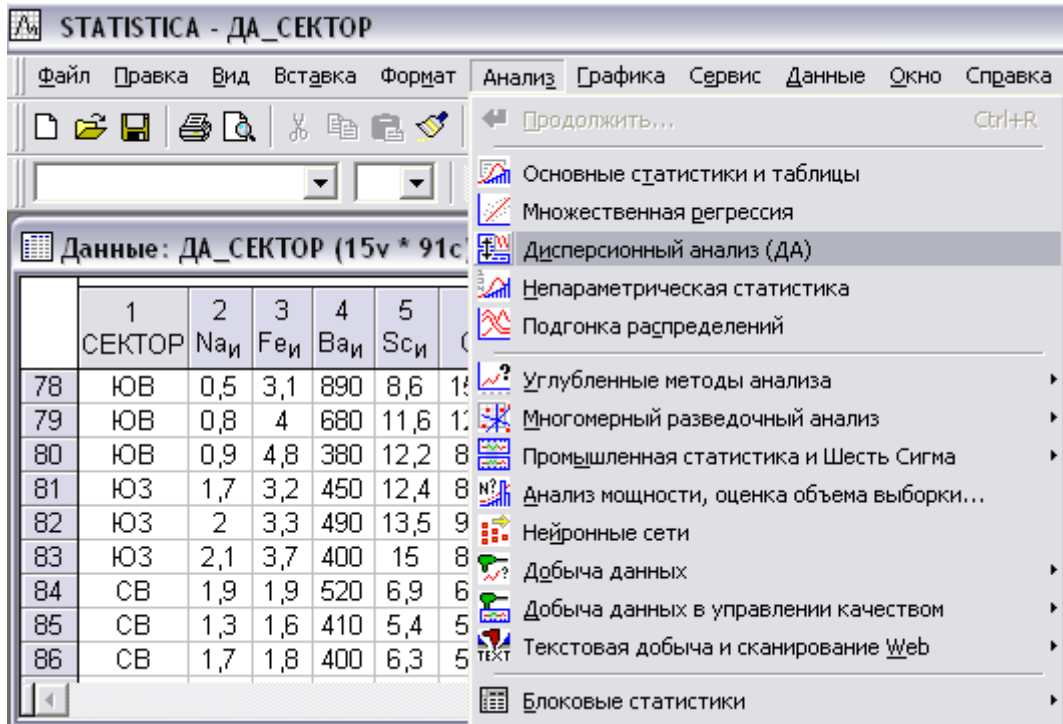


Рис.2.34. Модуль Дисперсионный анализ (ДА) в головном модуле Анализ с фрагментом базы данных ДА_СЕКТОР

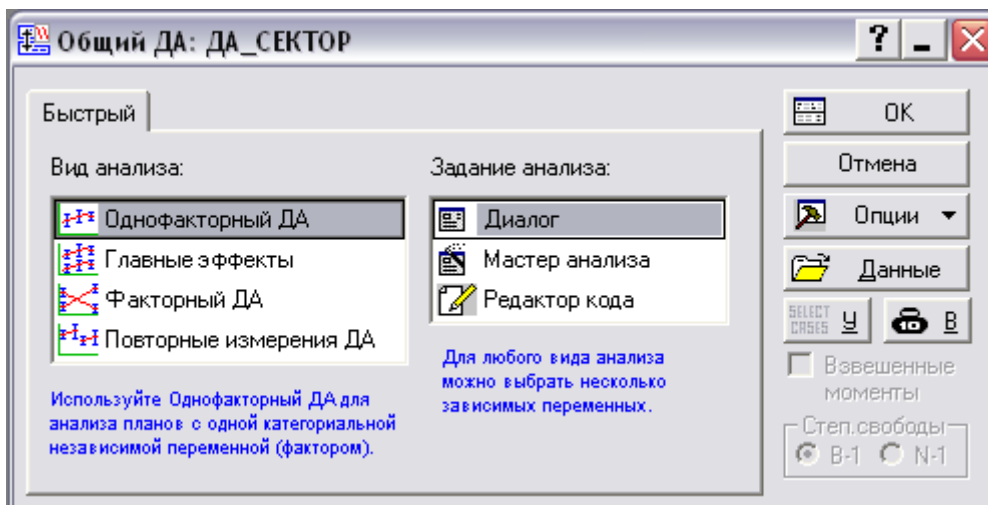


Рис.2.35. Выбор вида (однофакторный) и задания (диалог) дисперсионного анализа

Шаг 3. В появившемся окне выберите в качестве зависимых все переменные (позиции 2-15), а в качестве категориальной СЕКТОР (позиция 1), закажите все коды (уровни) фактора, как показано на рис.2.36, и нажмите ОК.

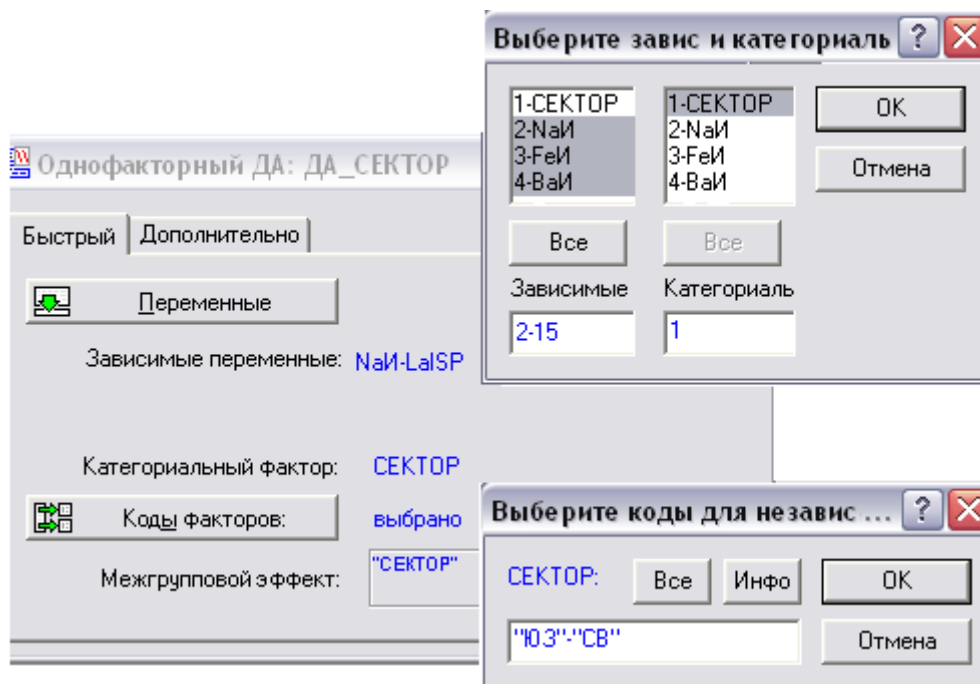


Рис. 2.36. Выбор переменных для дисперсионного анализа ДА СЕКТОР

Шаг 4. После этого в появившемся окне **Результаты анализа** (рис.2.37) в режиме **Итоги** нажмите **Статистики ячеек** для просмотра результатов описательной статистики (рис.2.38) и **Одномерн. результаты** для просмотра результатов дисперсионного анализа (рис.2.39).

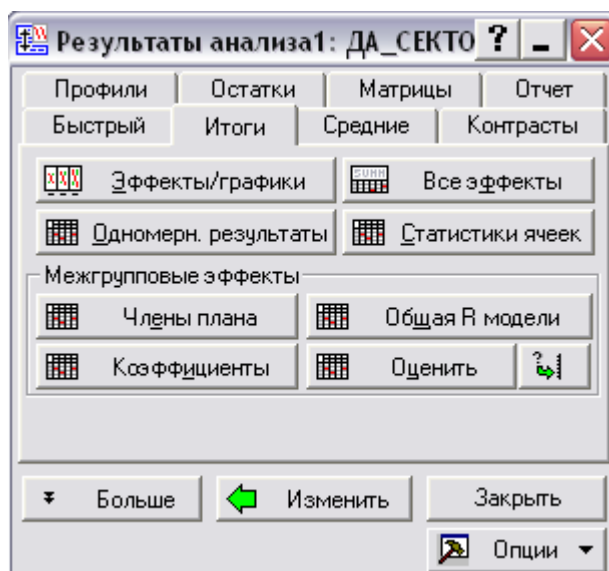


Рис. 2.37. Окно **Результаты анализа**

Эффект	Описательные статистики (ДА_СЕКТОР)							
	Уровень Фактор	N	Na _и Среднее	Na _и Ст. откл.	Na _и Стд. ош.	Na _и -95,00%	Na _и +95,00%	Fe _и Среднее
Всего		91	1,333	0,418	0,044	1,246	1,420	3,196
СЕКТОР	ЮЗ	22	1,786	0,253	0,054	1,674	1,899	3,577
СЕКТОР	ЮВ	61	1,115	0,282	0,036	1,042	1,187	3,134
СЕКТОР	СВ	8	1,750	0,334	0,118	1,471	2,029	2,612

Рис. 2.38. Результаты описательной статистики ДА_СЕКТОР


На рис. 2.38 кроме общей и групповых средних приведены также соответствующие стандартные отклонения, стандартные ошибки и границы 95% - ых доверительных интервалов по каждой переменной.





Эффект	Одномерные результаты для каждой ЗП (ДА_СЕКТОР)									
	Сигма-ограниченная параметризация Декомпозиция гипотезы									
	Степени свободы	Na _и SS	Na _и MS	Na _и F	Na _и p	Fe _и SS	Fe _и MS	Fe _и F	Fe _и p	
СЕКТОР	2	8,82	4,41	56,20	0,0000	6,15	3,08	7,73	0,0008	
Ошибка	88	6,90	0,08			35,01	0,40			
Всего	90	15,72				41,16				

Рис. 2.39. Результаты дисперсионного анализа ДА_СЕКТОР

На рис. 2.39 приведена таблица результатов дисперсионного анализа по каждой переменной. В столбце SS приведены значения сумм квадратов отклонений: например, для переменной Na_и SS_B = 8,818 – межгрупповой, характеризующей изменчивость между уровнями фактора СЕКТОР, SS_R = 6,904 – внутригрупповой, характеризующей изменчивость внутри уровней, и SS_T = 15,723 – общей вариацией переменной, складывающейся в силу (1.30) из двух компонент SS_B и SS_R. В первом столбце приведены соответствующие степени свободы: $\nu_B = 2$, $\nu_R = 88$ и $\nu_T = 90$. В столбце MS приведены значения средних квадратов, которые получают делением сумм квадратов отклонений на соответствующее число степеней свободы. В четвертом столбце приведено значение наблюдаемого F-отношения, вычисленное по формуле $F_{набл} = MS_B / MS_R$, а в последнем столбце - соответствующее значение уровня значимости p , определяющее значимость различий групповых средних по следующей качественной шкале: незначимо ($0,10 < p$), слабо значимо ($0,05 < p <$

0,10), статистически значимо ($0,005 < p < 0,05$), сильно значимо ($0,0005 < p < 0,005$) и высоко значимо ($p < 0,0005$). Полученный результат (рис. 2.39) является высоко значимым.

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

Шаг 4'. Для просмотра результатов дисперсионного анализа можно также нажать  Эффекты/графики (рис. 2.37). После этого в появившемся окне  Таблица всех эффектов (рис.2.40) выделить опцию  График и два раза щелкнуть на строке результатов для просмотра графика средних выбранной переменной, например, Na_{II} с усами 95% - ых доверительных интервалов (рис.2.41) или  Таблицу - для просмотра описательной статистики (рис. 2.38)

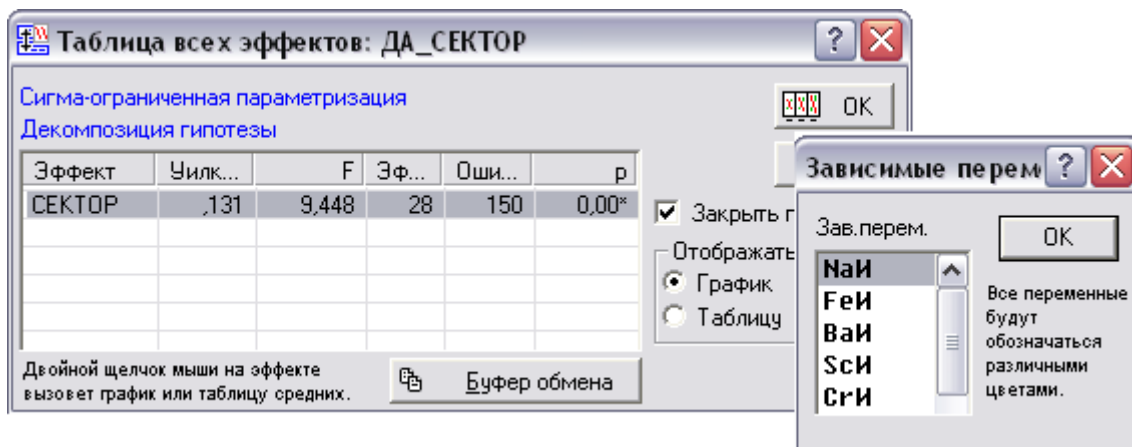


Рис. 2.40. Диалоговое окно  Таблица всех эффектов (содержит результаты анализа) для просмотра таблиц и графика средних

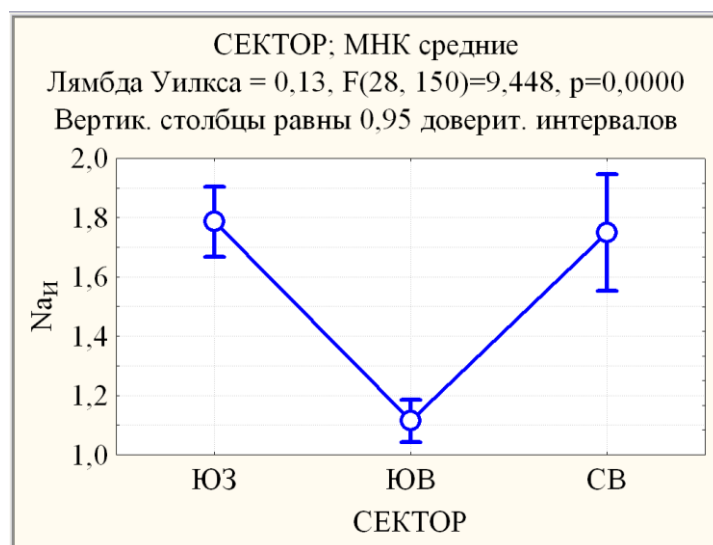

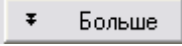
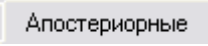
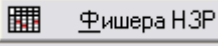
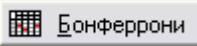

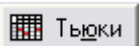
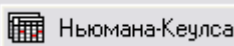
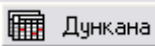


Рис. 2.41. График средних Na_{II} с усами 95% доверительных интервалов

Согласно многомерному Λ -критерию Уилкса (рис. 2.40 или подзаголовок рис. 2.41) фактор СЕКТОР по совокупности всех переменных высоко значим ($p = 0,0000 < 0,0005$). Поэтому далее для каждой переменной исследуется одномерный F -критерий (рис.2.39). Так, например, фактор СЕКТОР высоко значим ($p = 0,0000 < 0,0005$) для Na_{II} и сильно значим ($0,0005 < p = 0,0008 < 0,005$) для Fe_{II} . В случае значимого значения одномерного F -критерия далее можно использовать метод множественных сравнений для выявления тех групп, которые вызвали данный эффект и максимально отличаются от других. В пакете Statistica 6.1 доступны несколько процедур множественного сравнения (Шеффе, Тьюки, множественного t - метода согласно п. 1.2.3, а также Бонферрони, Ньюмана-Кеулса, Дункана) в которых за основу взяты различные распределения, что приводит, как правило, к разным доверительным интервалам.

Шаг 5. В окне  Результаты анализа (рис. 2.37) нажмите кнопку  и в открывшемся окне выберите режим , в котором представлены различные апостериорные критерии (рис. 2.42). Выберите последовательно все критерии , , , ,  и  (рис. 2.43).

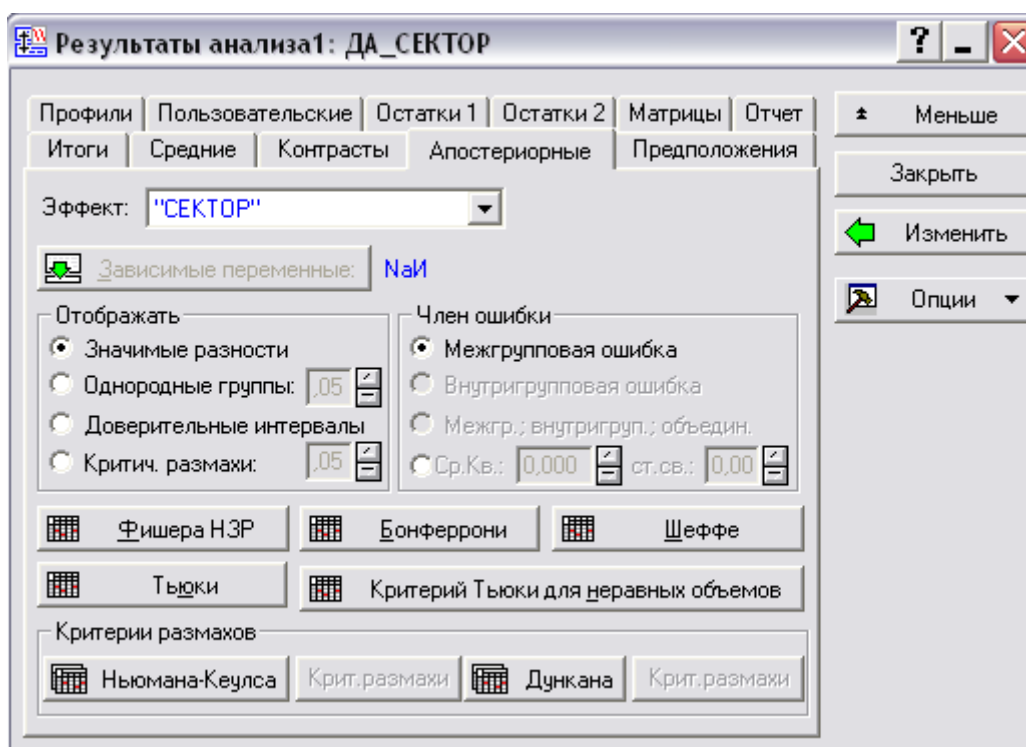
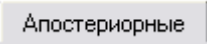


Рис. 2.42. Режим  для выбора апостериорных критериев множественных сравнений средних

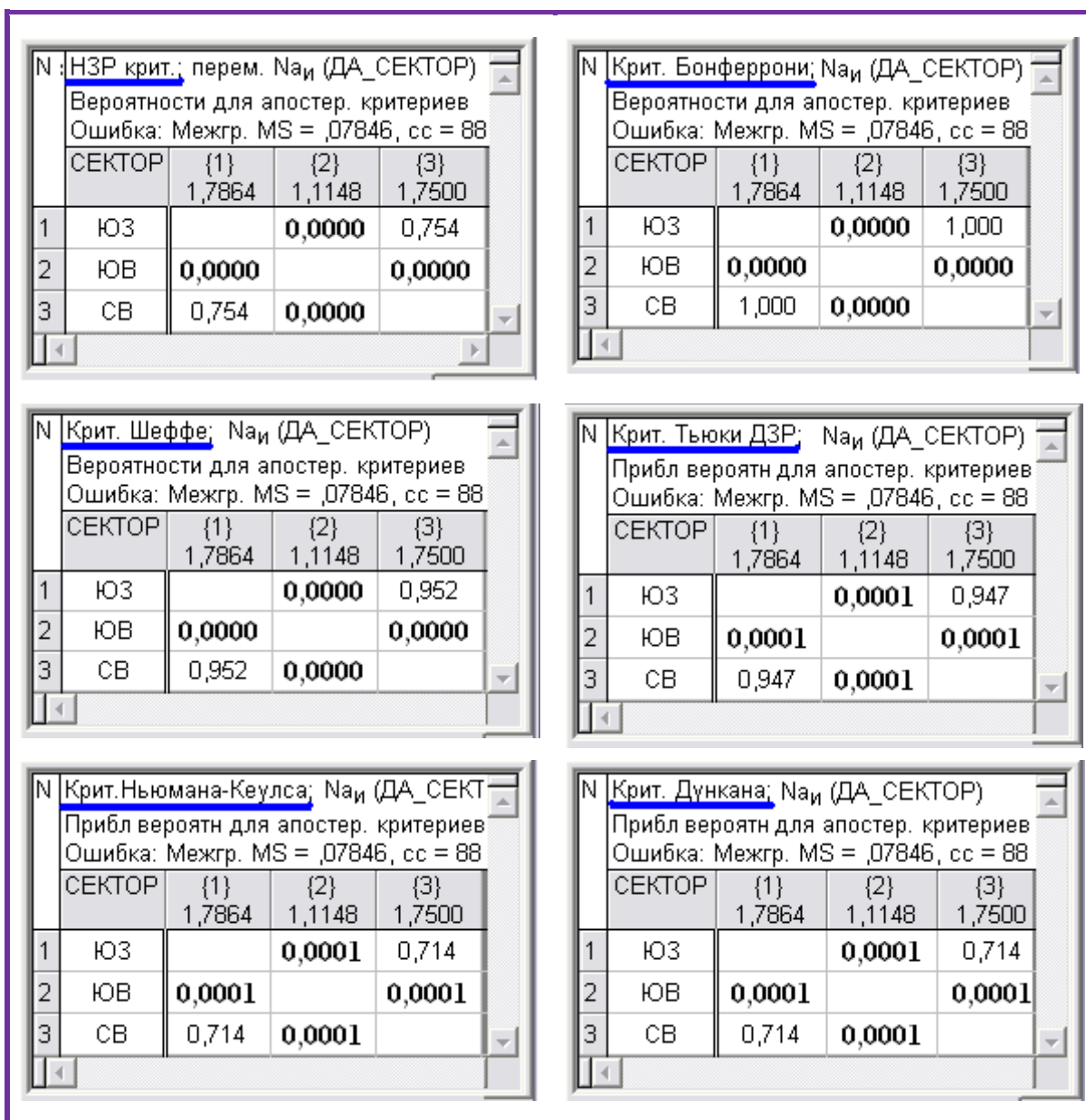


Рис. 2.43. Результаты сравнений средних групп $N_{и}$ разными апостериорными критериями

Согласно рис. 2.43, все апостериорные критерии единодушны: средние ЮЗ и СВ различаются незначимо ($0,10 < p$), но от ЮВ отличаются высоко значимо ($p < 0,0005$).

В общем случае различия в результатах применения разных критериев могут оказаться значимыми. Каждый из них имеет свои особенности и рекомендации по 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 иссле-

дования данных, в то время как в двух других методах множество контрастов может быть любым.

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

Шаг 6. В окне **Результаты анализа** (рис. 2.42) выберите режим **Предположения**, в котором представлены различные критерии проверки гипотезы однородности дисперсий и графические средства проверки соответствия закона распределения переменной нормальному закону (рис. 2.44). Выберите, например, критерий однородности дисперсий **Критерий Левена** (рис. 2.45) и **Гистограммы** распределения переменной для всех групп (рис. 2.46).

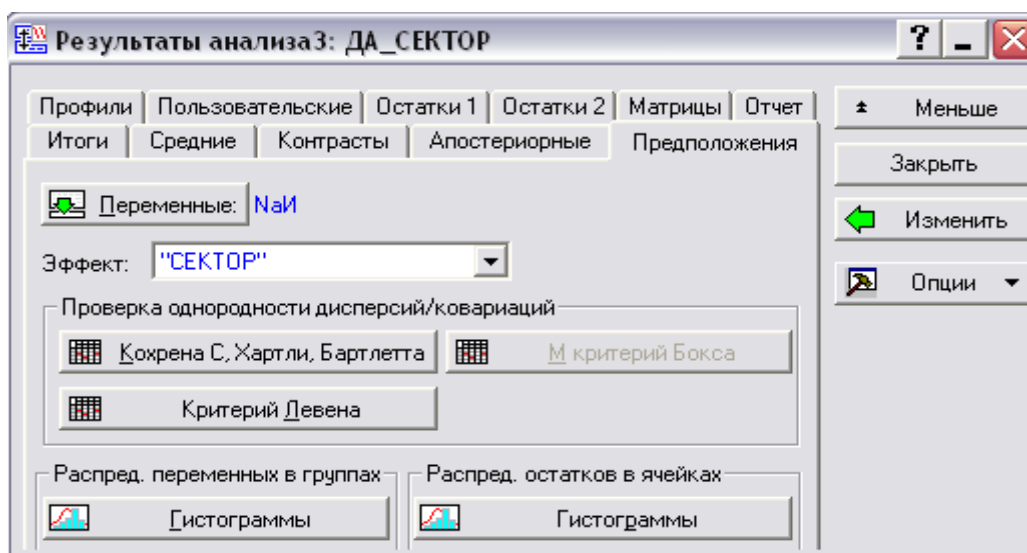


Рис. 2.44. Режим **Предположения** для проверки предположений дисперсионного анализа

Критерий Левена однородности дисперсий				
Эффект: "СЕКТОР"				
Степени свободы для всех F: 2, 88				
	MS	MS	F	p
Эффект	Ошибка			
NaИ	0,012	0,027	0,463	0,631

Рис. 2.45. Результаты проверки гипотезы об однородности дисперсий с помощью **Критерий Левена**

Согласно рис. 2.45 во всех группах дисперсии различаются незначимо ($0,10 < p$).

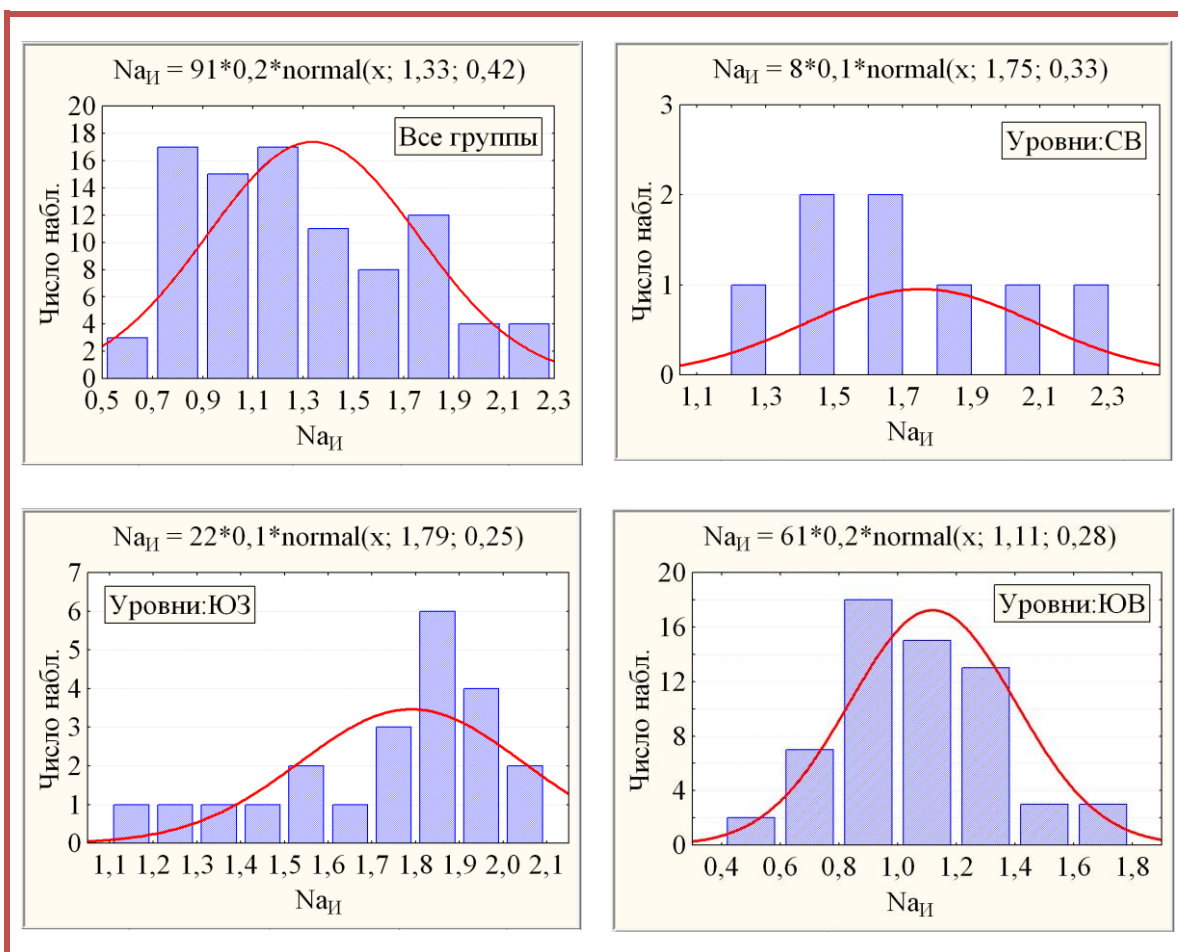


Рис. 2.46. Гистограммы распределений переменной для всех групп

Согласно рис. 2.46, распределение групп уровней фактора СЕКТОР зримо отличаются от нормального закона (см. также рис. 2.12 и 2.14).

В связи с нарушением предположений о нормальном распределении сравниваемых групп, соответствующих уровням фактора СЕКТОР, и об однородности дисперсий в группах обратимся к непараметрическим альтернативам однофакторного дисперсионного анализа: тесту Краскела-Уоллиса и медианному тесту.

Шаг 7. Запустите модуль **Анализ** и щелкните в нем по кнопке **Непараметрическая статистика** (см. рис.2.34). Затем выберите сравнение нескольких независимых групп **Сравнение нескольких независимых групп** (рис.2.47), закажите переменные стандартным образом (см. рис. 2.36) и щелкните по кнопке **Да Краскела-Уоллиса и медианный тест** (рис.2.48). В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа: по тесту Краскела-Уоллиса (рис.2.49) и медианному тесту (см. рис.2.50).

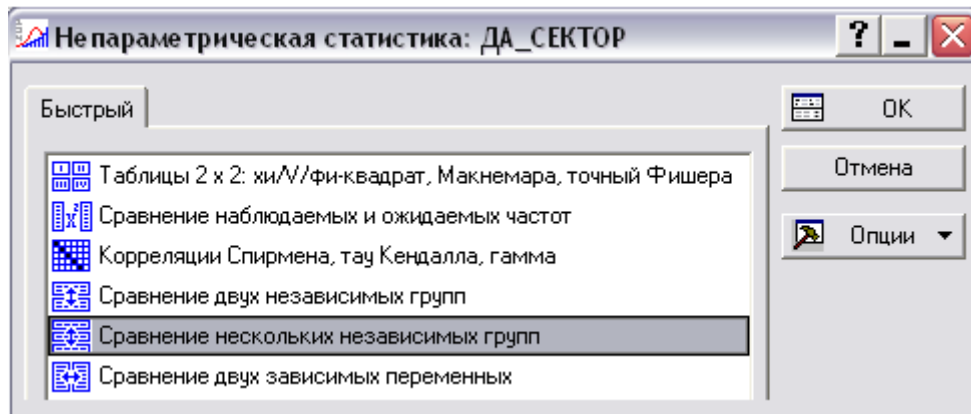


Рис.2.47. Модуль Не параметрическая статистика.

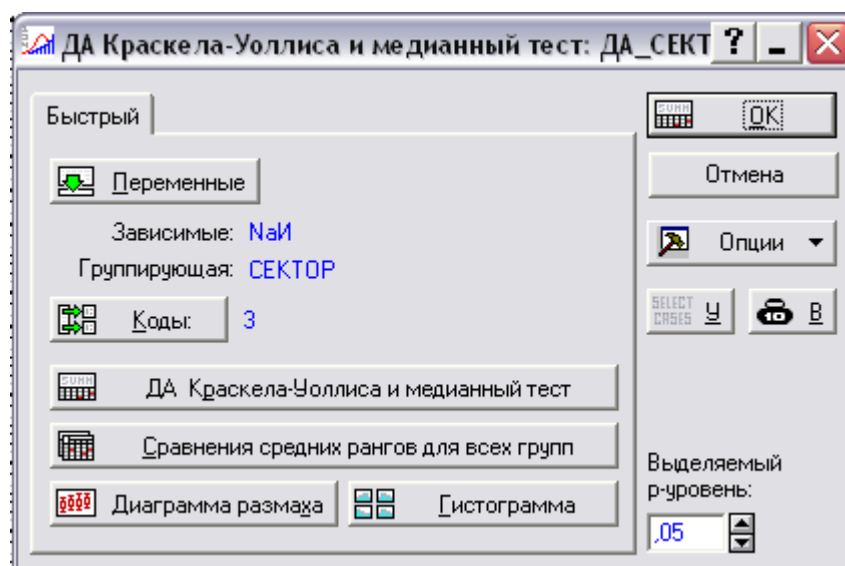


Рис.2.48. Диалоговое окно выбора переменных и тестов непараметрического однофакторного дисперсионного анализа

Зависим.: NaИ	Ранговый ДА Краскела-Уоллиса; NaИ (ДА_СЕКТОР)		
	Группирующая переменная: СЕКТОР		
	Кр.Краскела-Уоллиса: $H(2, N=91) = 47,93252$ $p = ,0000$		
	Код	Допуст N	Сумма Ряды
ЮЗ	105	22	1627,500
ЮВ	106	61	1988,500
СВ	107	8	570,000

Рис.2.49. Результаты рангового дисперсионного анализа СЕКТОР по тесту Краскела-Уоллиса


Зависимые: Na _и	Медианный тест, общ. медиана = 1,30000; Na _и (ДА_СЕКТОР)			
	Группирующая переменная: СЕКТОР Хи-квадрат = 40,64219 сс = 2 p = ,0000			
	ЮЗ	ЮВ	СВ	Всего
<= Медиана: наблюд.	2,00	49,00	1,00	52,00
ожидаемые	12,57	34,86	4,57	
набл.-ожд.	-10,57	14,14	-3,57	
> Медиана: наблюд.	20,00	12,00	7,00	39,00
ожидаемые	9,43	26,14	3,43	
набл.-ожд.	10,57	-14,14	3,57	
Сумма: наблюд.	22,00	61,00	8,00	91,00

Рис.2.50. Результаты рангового дисперсионного анализа СЕКТОР по медианному тесту

Согласно критерию Краскела-Уоллиса (рис. 2.49) и медианному тесту (рис. 2.50) существуют высоко значимые ($p < 0,0005$) различия между средними в группах, что подтверждает результат параметрического дисперсионного анализа о высоко значимой территориальной неоднородности распределения Na_и по совокупности секторов.

Так как, согласно результатам дисперсионного анализа (рис.2.49), существуют значимые различия между средними в группах, то далее можно использовать непараметрический аналог метода множественных сравнений.

Шаг 8. В окне **ДА Краскела-Уоллиса** (рис. 2.48) нажмите кнопку

 Сравнения средних рангов для всех групп (рис. 2.51).

Зависим.: Na _и	Сравнения p значений (2-стороннее); Na _и (ДА_СЕКТОР)		
	Группирующая переменная: СЕКТОР Кр.Краскела-Уоллиса: H (2, N= 91) =47,93252 p =,0000		
	ЮЗ R:73,977	ЮВ R:32,598	СВ R:71,250
ЮЗ		0,00000	1,000
ЮВ	0,00000		0,00030
СВ	1,000	0,00030	

Рис. 2.51. Результаты сравнений средних рангов групп

Согласно рис. 2.51, средние ЮЗ и СВ различаются незначимо ($0,10 < p$), но от ЮВ отличаются высоко значимые ($p < 0,0005$), что также подтверждает результат параметрического дисперсионного ана-

лиза о значимой территориальной неоднородности распределения N_{a_i} по секторам.

Шаг 8'. В окне **ДА Краске ла-Уоллиса** (рис.2.48) нажмите кнопку **Диagramма размаха** для просмотра диаграммы размаха по группам (рис. 2.52), аналогично рис. 2.41, и – **Гистограмма** для просмотра гистограмм по группам (рис.2.46)

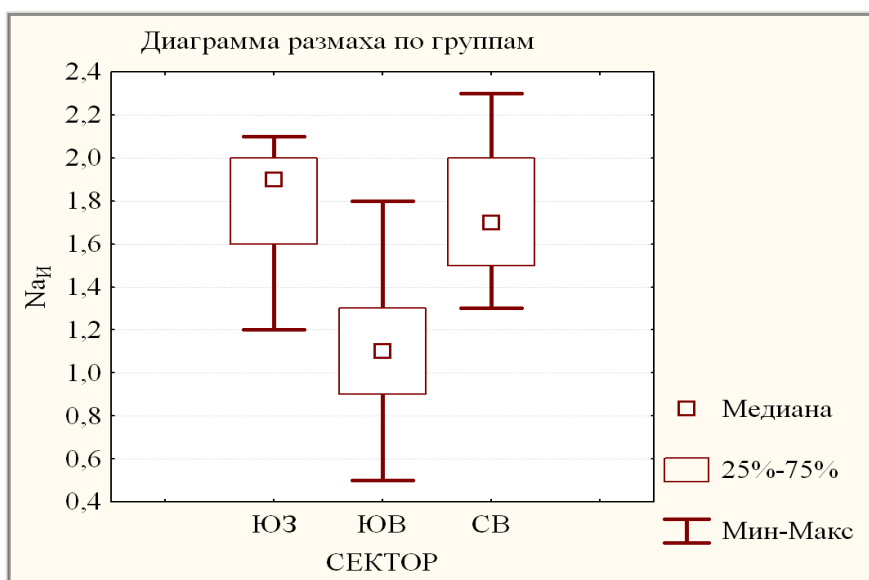


Рис. 2.52. Диаграммы размаха по группам (квадрат – медиана, прямоугольник – квартильных размах, усы – абсолютный размах)

В случае расхождения результатов непараметрического и параметрического однофакторного дисперсионного анализа в связи с нарушением условий применения последнего (предположений о нормальном распределении сравниваемых групп и об однородности дисперсий в группах) более корректными признаются результаты непараметрического дисперсионного анализа.

2.3.2. Двухфакторный дисперсионный анализ

Рассмотрим далее применение двухфакторного дисперсионного анализа с фиксированными эффектами к примеру содержания химических элементов для проведения оценки значимости совместного влияния факторов МЕТОД и СЕКТОР на переменную N_a . На предварительном этапе создается в STATISTICA файл (с названием, напр., ДА МЕТОД-СЕКТОР.sta) с категориальной базой данных (рис.2.53). При этом названия факторов и переменных выносятся в названия столбцов.

	1 СЕКТОР	2 МЕТОД	3 Na	4 Fe	5 Ba	6 Sc	7 Cr	8 Co	9 La
89	СВ	ИНАА	2,3	3,1	390	12,8	88,3	13,9	31,7
90	СВ	ИНАА	1,7	3,4	370	14,2	84,1	15,1	33,4
91	СВ	ИНАА	1,5	3,3	400	15,5	175	16,6	27,5
92	ЮЗ	ISP	2,16	3,16	370	10,2	144,5	21,1	28
93	ЮЗ	ISP	2,27	3,22	350	9,1	161,2	21,6	26,2
94	ЮЗ	ISP	2,32	3,61	330	10	127,7	23	24,9

Рис.2.53. Категориальная база данных ДА МЕТОД-СЕКТОР

Аналогично случаю фактора СЕКТОР (рис.2.34) на основе базы данных (см. Приложения П₁), в данном случае кроме фактора МЕТОД (с двумя уровнями ИНАА и ISP) учтен в рассмотрении второй фактор СЕКТОР с тремя уровнями: юго-западный (ЮЗ) сектор (измерения в П₁ №1–19 и №81–83), юго-восточный (ЮВ) сектор (измерения №20–80) и северо-восточный (СВ) сектор (измерения №84–91).

Шаг 1. Запустите модуль **Анализ** и щелкните в нем по кнопке **Дисперсионный анализ (ДА)** (аналогично рис.2.34).

Шаг 2. В появившемся окне выберите тип анализа (многофакторный дисперсионный анализ) и задайте метод (диалог быстрых спецификаций), как показано на рис.2.54, и нажмите **OK**.

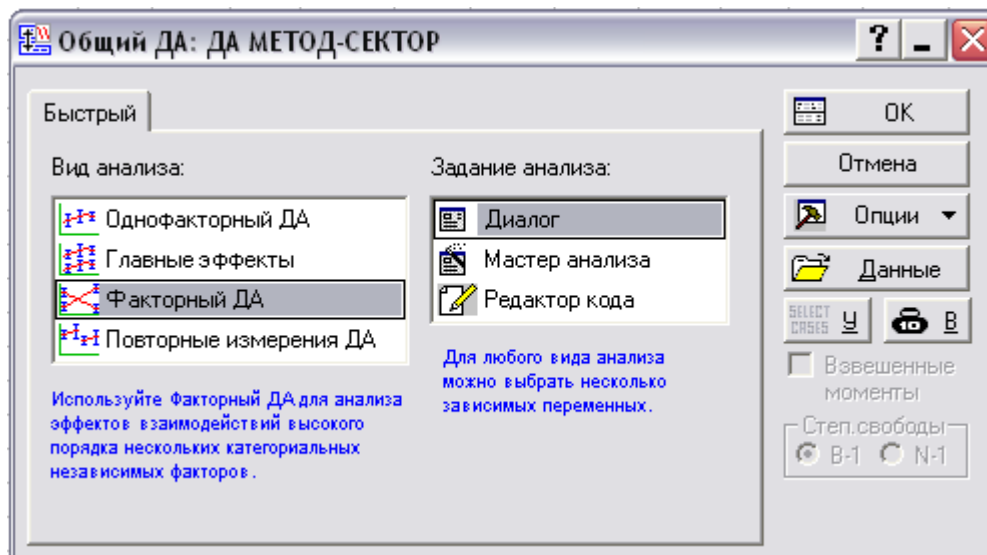


Рис.2.54. Выбор типа и метода ДА МЕТОД-СЕКТОР

Шаг 3. В появившемся окне выберите переменные, как показано на рис.2.55, и нажмите **OK**.

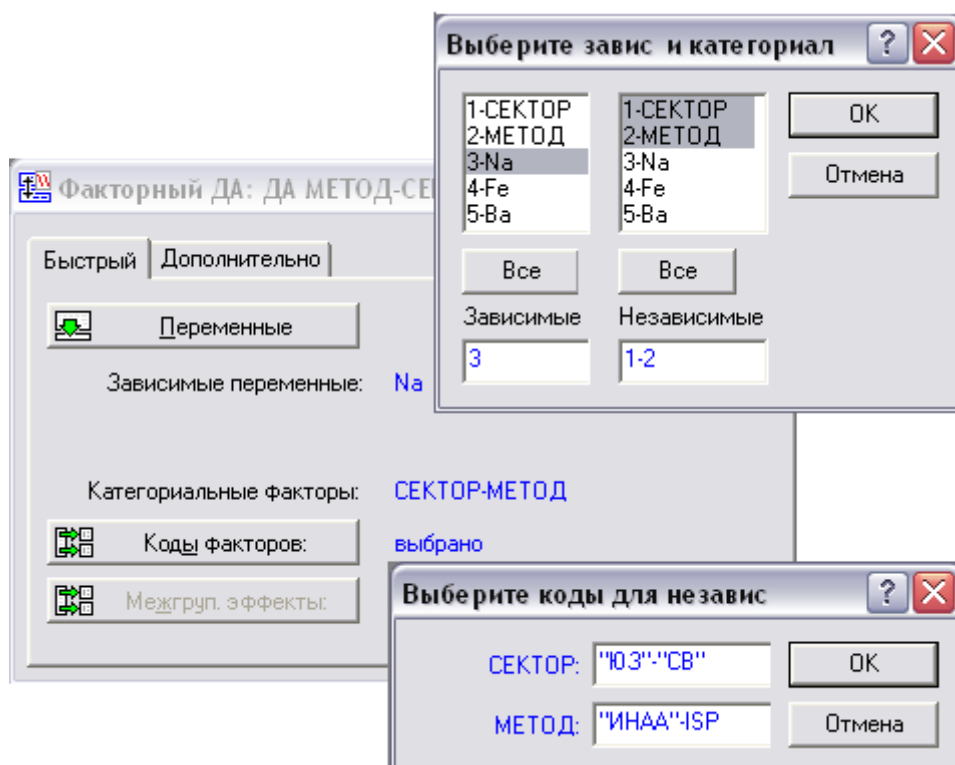


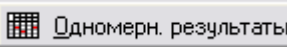



Рис. 2.55. Выбор переменных для ДА МЕТОД-СЕКТОР

Шаг 4. После этого в появившемся окне  Результаты анализа (рис.2.37) в режиме  Итоги нажмите  Одномерн. результаты для просмотра результатов дисперсионного анализа (рис.2.56) и  Статистики ячеек для просмотра результатов описательной статистики (рис.2.57).

Эффект	Одномерные результаты для каждой ЗП (ДА МЕТОД-СЕКТОР) Сигма-ограниченная параметризация Декомпозиция гипотезы				
	Степени свободы	Na SS	Na MS	Na F	Na p
СЕКТОР	2	18,616	9,308	99,248	0,000000
МЕТОД	1	2,358	2,358	25,141	0,000001
СЕКТОР*МЕТОД	2	1,889	0,945	10,073	0,000072
Ошибка	176	16,506	0,094		
Всего	181	46,341			

Рис. 2.56. Результаты дисперсионного анализа МЕТОД-СЕКТОР

Таблица, изображенная на рис.2.56, объединяет главные результаты анализа. Можно заметить, что все эффекты (каждый фактор МЕТОД и СЕКТОР, а также их взаимодействие МЕТОД*СЕКТОР) являются высоко значимыми ($p < 0,0005$).

Эффект	Описательные статистики (ДА МЕТОД-СЕКТОР)							
	Уровень Фактор	Уровень Фактор	N	На Среднее	На Ст.откл.	На Стд.ош.	На -95,00%	На +95,00%
Всего			182	1,559	0,506	0,038	1,485	1,633
СЕКТОР	ЮЗ		44	2,103	0,449	0,068	1,967	2,240
СЕКТОР	ЮВ		122	1,349	0,375	0,034	1,282	1,416
СЕКТОР	СВ		16	1,668	0,367	0,092	1,472	1,864
МЕТОД	ИНАА		91	1,333	0,418	0,044	1,246	1,420
МЕТОД	ISP		91	1,786	0,486	0,051	1,685	1,887
СЕКТОР*МЕТОД	ЮЗ	ИНАА	22	1,786	0,253	0,054	1,674	1,899
СЕКТОР*МЕТОД	ЮЗ	ISP	22	2,420	0,371	0,079	2,256	2,585
СЕКТОР*МЕТОД	ЮВ	ИНАА	61	1,115	0,282	0,036	1,042	1,187
СЕКТОР*МЕТОД	ЮВ	ISP	61	1,583	0,305	0,039	1,505	1,661
СЕКТОР*МЕТОД	СВ	ИНАА	8	1,750	0,334	0,118	1,471	2,029
СЕКТОР*МЕТОД	СВ	ISP	8	1,586	0,403	0,143	1,249	1,923

Рис. 2.57. Результаты описательной статистики ДА МЕТОД-СЕКТОР

Шаг 4'. Графики средних разных эффектов (рис.2.56) можно построить аналогично рис.2.41 или в режиме **Средние** (рис.2.37), выбирая нужный эффект и нажимая кнопку **График** (рис.2.58). Перебирая разные эффекты, можно построить графики средних фактора МЕТОД (рис.2.59), фактора СЕКТОР (рис.2.60) и взаимодействия факторов МЕТОД и СЕКТОР (рис.2.61 или 2.62).

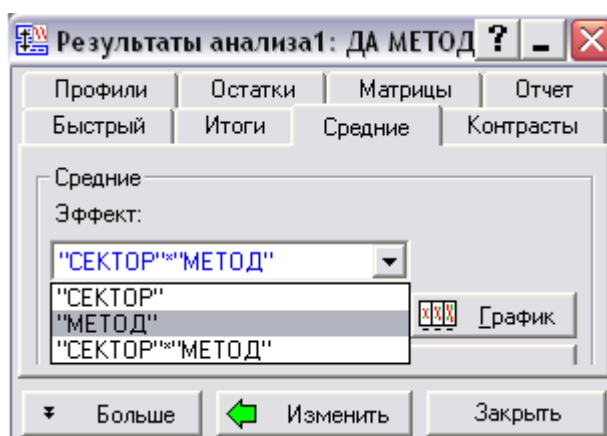


Рис. 2.58. Режим **Средние** диалогового окна **Результаты анализа**

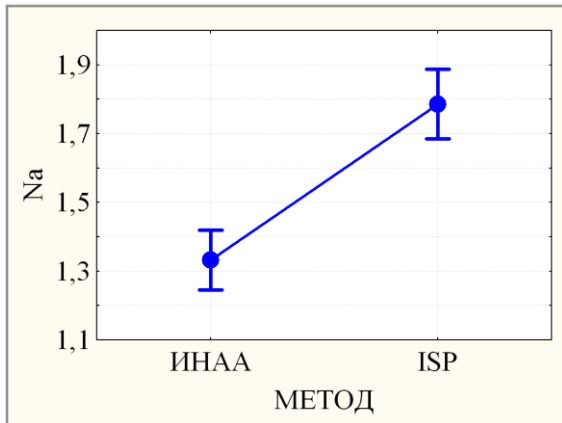


Рис. 2.59. График средних МЕТОД

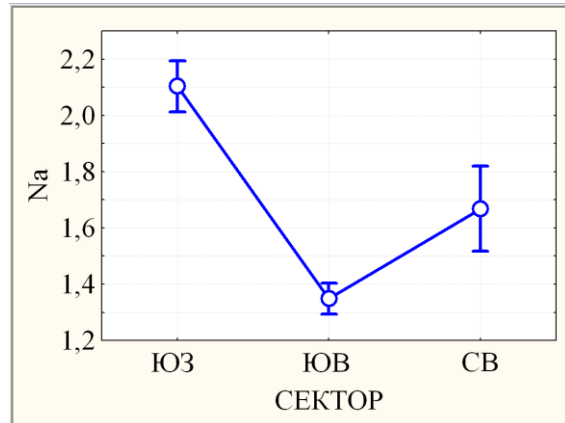


Рис. 2.60. График средних СЕКТОР

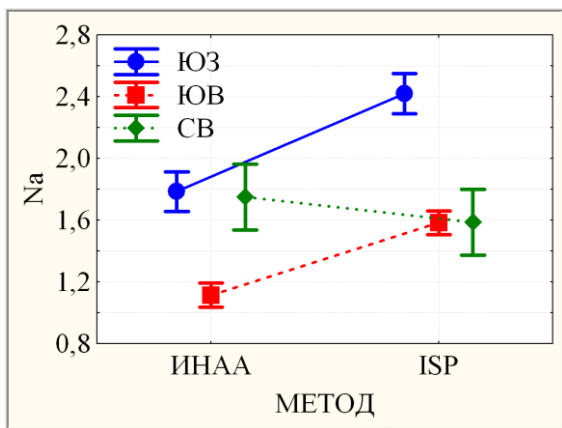


Рис. 2.61. Графики средних СЕКТОР*МЕТОД

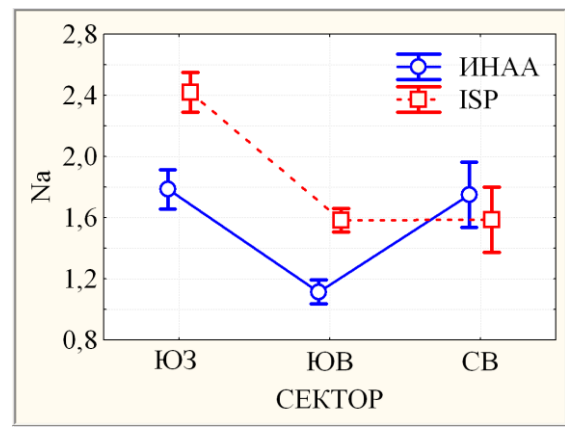


Рис. 2.62. Графики средних МЕТОД*СЕКТОР

Заметим, что на рис. 2.60 изображен график средних фактора СЕКТОР, усредненных по методам ИНАА и ISP, в отличие от графика средних фактора СЕКТОР по методу ИНАА рис. 2.41.

Сравнение рис. 2.59 с рис.2.61 или рис. 2.60 с рис.2.62 наглядно демонстрирует эффекты взаимодействия между факторами. Согласно рис. 2.59 среднее Na_{ISP} больше среднего $Na_{ИНАА}$, но с учетом фактора СЕКТОР выясняется, что это верно только для двух южных секторов (ЮЗ и ЮВ), а для северного – наоборот. Тот же вывод следует из сравнения рис. 2.60 с рис.2.62.

Далее применим метод множественных сравнений.

Шаг 5. В окне Результаты анализа (2.58) нажмите кнопку Больше и в открывшемся окне выберите режим Апостериорные, в котором представлены различные апостериорные критерии (рис.2.63). Далее выберите эффект СЕКТОР*МЕТОД. Затем закажите поочередно режимы отображения «Значимые различия» (рис.2.64) и «Однородные группы» (рис.2.65) для критерия Фишера НЗР.

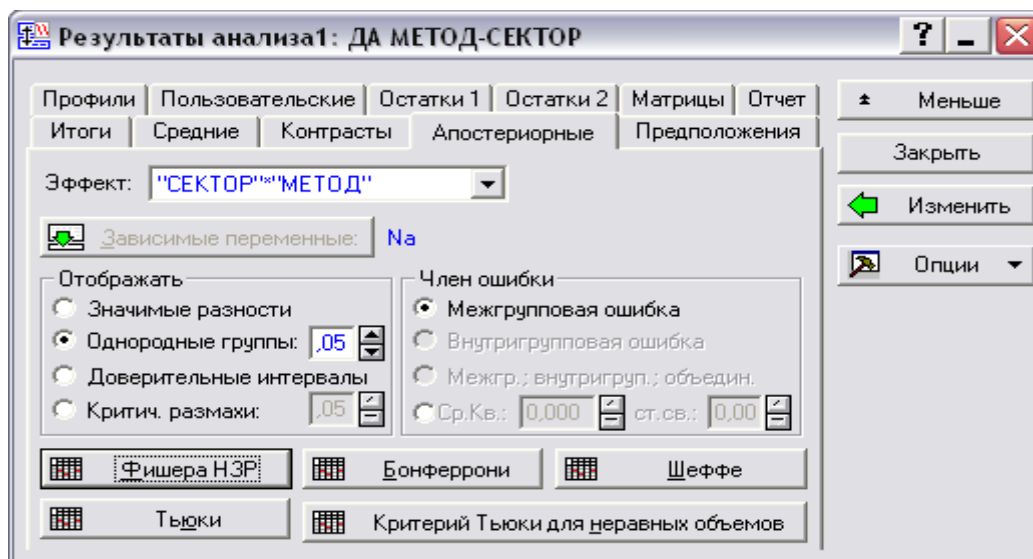


Рис. 2.63. Режим **Апостериорные** для выбора апостериорных критериев множественных сравнений средних

N НЗР крит.; перем. Na (ДА МЕТОД-СЕКТОР)
 Вероятности для апостер. критериев
 Ошибка: Межгр. MS = ,09378, сс = 176,00

	СЕКТОР	МЕТОД	{1}	{2}	{3}	{4}	{5}	{6}
			1,7864	2,4205	1,1148	1,5831	1,7500	1,5863
1	ЮЗ	ИНАА		0,00000	0,00000	0,00832	0,774	0,115
2	ЮЗ	ISP	0,00000		0,00000	0,00000	0,00000	0,00000
3	ЮВ	ИНАА	0,00000	0,00000		0,00000	0,00000	0,00006
4	ЮВ	ISP	0,00832	0,00000	0,00000		0,149	0,978
5	СВ	ИНАА	0,774	0,00000	0,00000	0,149		0,286
6	СВ	ISP	0,115	0,00000	0,00006	0,978	0,286	

Рис. 2.64. Результаты НЗР в режиме «Значимые различия»

N ячеек НЗР крит.; перем. Na (ДА МЕТОД-СЕКТОР)
 Однородные группы, alpha = ,05000
 Ошибка: Межгр. MS = ,09378, сс = 176,00


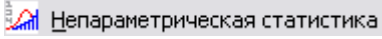
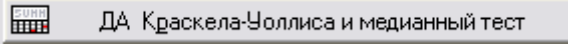
	СЕКТОР	МЕТОД	Na	1	2	3	4
			Среднее				
3	ЮВ	ИНАА	1,114754			****	
4	ЮВ	ISP	1,583089	****			
6	СВ	ISP	1,586250	****	****		
5	СВ	ИНАА	1,750000	****	****		
1	ЮЗ	ИНАА	1,786364		****		
2	ЮЗ	ISP	2,420455				****

Рис. 2.65. Результаты НЗР в режиме «Однородные группы»

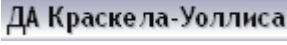

Согласно рис.2.64-65, для эффекта взаимодействия СЕКТОР*МЕТОД можно выделить однородные (различающиеся незначимо, то есть на уровне значимости $p > 0,10$) кластеры групп, расположенные в порядке убывания средних: {ЮЗ*ISP}, {ЮЗ*ИНАА, СВ*ИНАА, СВ*ISP}, {СВ*ИНАА, СВ*ISP, ЮВ*ISP}, {ЮВ*ИНАА}. Имеются два пересекающихся кластера так, что ЮЗ*ИНАА отличается от ЮВ*ISP статистически значимо (на уровне $0,005 < p \approx 0,008 < 0,05$).

Шаг 6. Для проверки предположений, лежащих в основе одномерного дисперсионного анализа выполните действия, аналогичные построению рис.2.44-2.46.

Продолжим анализ данных с использованием непараметрических методов (тест Краскела-Уоллиса и медианный тест). В связи с этим построим категориальную переменную С*М из 6-ти уровней взаимодействия (ЮЗ*ИН, ЮЗ*IS, СВ*ИН, СВ*IS, ЮВ*ИН, ЮВ*IS) и обратимся к непараметрическим альтернативам однофакторного дисперсионного анализа (аналогично рис. 2.47 – 2.52).

Шаг 7. Запустите модуль , щелкните в нем по кнопке  и выберите множественные сравнения групп (рис.2.47). Закажите переменные стандартным образом, беря в качестве категориальной переменной С*М, а зависимой – Na, и щелкните по кнопке  (рис. 2.48).

В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа (аналогично рис. 2.49 – 2.50). Согласно критерию Краскела-Уоллиса и медианному тесту существуют высоко значимые ($p < 0,0005$) различия между средними в группах взаимодействий. Это подтверждает результат параметрического дисперсионного анализа о высоко значимой неоднородности распределения Na по совокупности уровней фактора взаимодействия СЕКТОР*МЕТОД и поэтому предполагает в дальнейшем использование непараметрического аналога метода множественных сравнений.

Шаг 8. В окне  (рис. 2.48) нажмите кнопку  (рис. 2.66).

Сравнивая рис. 2.66 с рис.2.64, можно заметить, что результаты непараметрического множественных сравнений в целом выглядят более мягкими, размывающими грани однородности кластеров групп С*М. Согласно рис.2.66, для эффекта взаимодействия СЕКТОР*МЕТОД можно выделить однородные кластеры групп, расположенные в порядке убывания средних рангов: {ЮЗ*ISP, ЮЗ*ИНАА, СВ*ИНАА}, {ЮЗ*ИНАА, СВ*ИНАА, СВ*ISP, ЮВ*ISP}, {СВ*ISP, ЮВ*ИНАА}.

Имеются три пересекающихся кластера так, что ЮЗ*ISP отличается от СВ*ISP статистически значимо (на уровне $0,005 < p \approx 0,038 < 0,05$), ЮЗ*ИНАА отличается высоко значимо (на уровне $p < 0,0005$) от ЮВ*ИНАА, которая в третьем кластере отличается от СВ*ISP слабо значимо (на уровне $0,05 < p \approx 0,06 < 0,10$).

Зависим.: Na	Сравнения p значений (2-стороннее); Na (ДА МЕТОД-СЕКТОР) Группирующая переменная: С*М Кр.Краскела-Уоллиса: H (5, N= 182) =108,7330 p =0,000					
	ЮЗ*ИН R:126,00	ЮВ*ИН R:42,066	СВ*ИН R:116,75	ЮЗ*ИС R:164,86	ЮВ*ИС R:97,721	СВ*ИС R:99,125
ЮЗ*ИН		0,00000	1,000	0,216	0,463	1,000
ЮВ*ИН	0,00000		0,00245	0,00000	0,00000	0,060
СВ*ИН	1,000	0,00245		0,404	1,000	1,000
ЮЗ*ИС	0,216	0,00000	0,404		0,00000	0,03762
ЮВ*ИС	0,463	0,00000	1,000	0,00000		1,000
СВ*ИС	1,000	0,060	1,000	0,03762	1,000	

Рис. 2.66. Результаты сравнений средних рангов групп С*М

Шаг 8'. В окне **ДА Краскела-Уоллиса** (рис.2.48) нажмите кнопку **Диаграмма размаха** для просмотра диаграммы размаха по группам (рис. 2.67), аналогично рис. 2.52, и **Гистограмма** для просмотра гистограмм по группам (рис. 2.68).

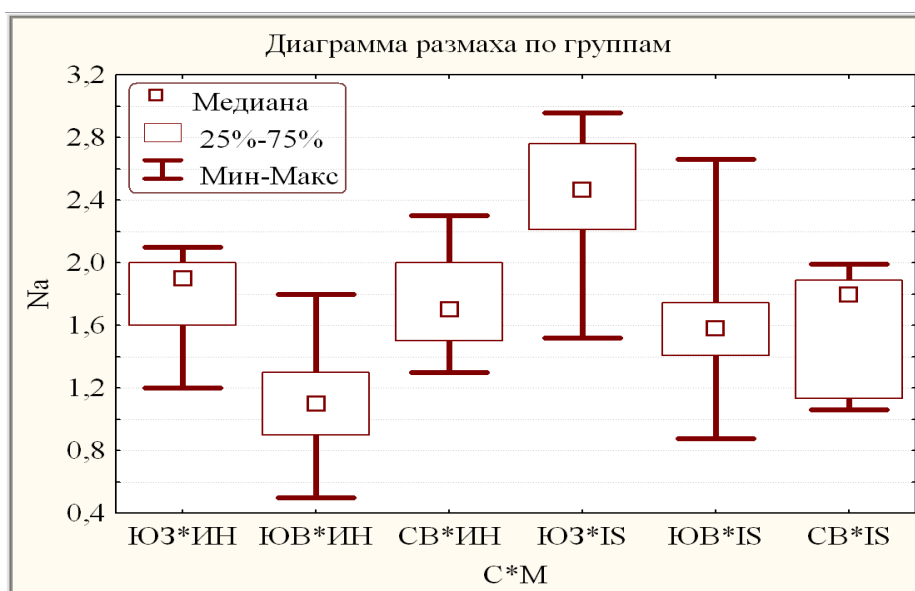


Рис. 2.67. Диаграммы размаха по группам С*М (квадрат – медиана, прямоугольник – квартильных размах, усы – абсолютный размах)

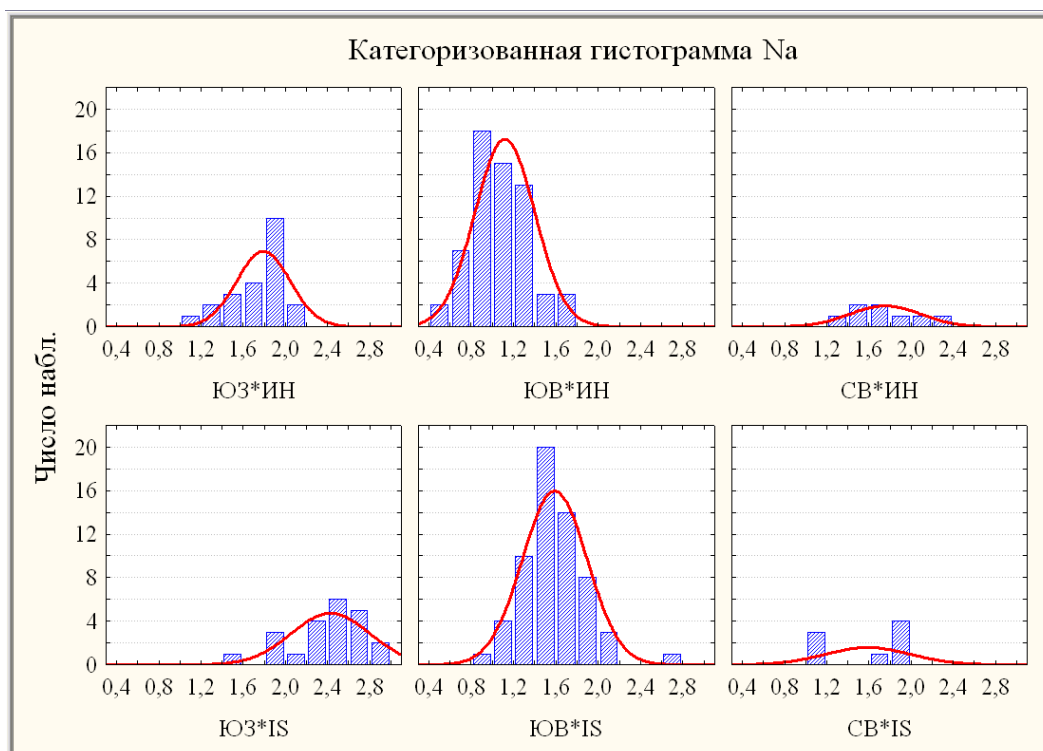


Рис. 2.68. Гистограммы распределения Na для всех групп С*М

2.4. Корреляционно – регрессионный анализ

2.4.1. Вычисление корреляционной матрицы ассоциации содержаний микроэлементов

В первой части этапа III для вычисления корреляционной матрицы одной из двух ассоциаций необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль **Анализ** (рис.2.2).

Шаг 2. Нажмите **Основные статистики и таблицы** и в появившемся окне

выберите **Парные и частные корреляции** и кликните **ОК**.

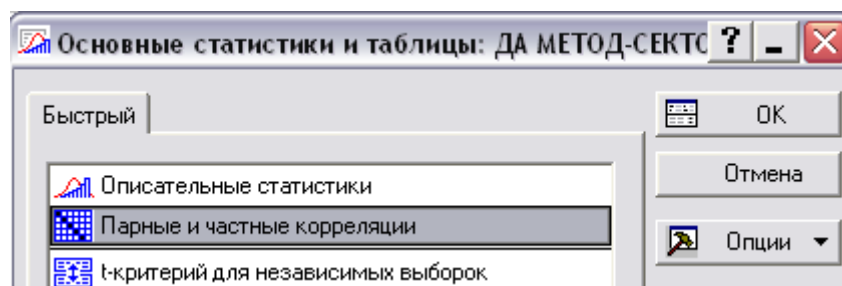


Рис. 2.69. Выбор **Парные и частные корреляции** в окне

Основные статистики и таблицы:

Шаг 4. В открывшемся окне **Парные и частные корреляции** нажмите левую кнопку **Квадратная матрица**, выберите все переменные метода ИНАА базы данных

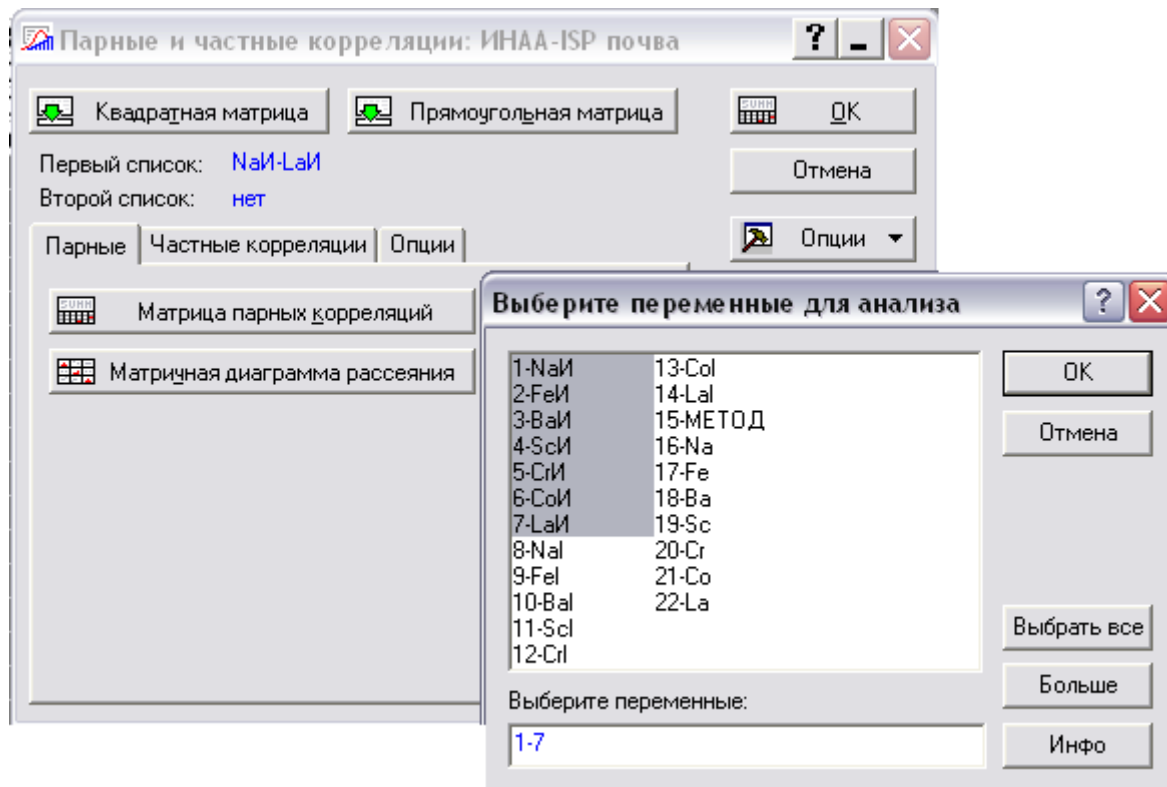


Рис. 2.70. Выбор микроэлементов (метод ИНАА) для корреляционного анализа.

Шаг 5. После этого нажмите кнопку **Матрица парных корреляций**.
Следующая таблица с результатами появится на экране:

Таблица 2.5

Корреляционная матрица выборки содержания микроэлементов в почво-грунтах, полученных методом ИНАА.

	Na _и	Fe _и	Ba _и	Sc _и	Cr _и	Co _и	La _и
Na _и	1,00	0,14	-0,35	0,37	0,17	0,03	0,45
Fe _и	0,14	1,00	0,11	0,87	0,24	0,38	0,64
Ba _и	-0,35	0,11	1,00	-0,09	0,13	0,33	-0,20
Sc _и	0,37	0,87	-0,09	1,00	0,18	0,24	0,76
Cr _и	0,17	0,24	0,13	0,18	1,00	0,59	0,13
Co _и	0,03	0,38	0,33	0,24	0,59	1,00	0,14
La _и	0,45	0,64	-0,20	0,76	0,13	0,14	1,00

Значимые, то есть отличные от нуля на уровне значимости 0,05, коэффициенты парной корреляции Пирсона выделены в таблице системой STATISTICA на экране красным цветом (в табл.2.5 нижняя критическая граница **0,24** значимых коэффициентов парной корреляции Пирсона выделена жирным шрифтом).

Аналогичным образом строится корреляционная матрица ассоциации микроэлементов по методу ISP (табл. 2.6)

В данном случае критическое значение (**0,24** в методе ИНАА и **0,22** в методе ISP) значимых, то есть отличных от нуля на уровне значимости 0,05, коэффициентов парной корреляции Пирсона определялась выбором наименьшего из значимых, выделенных в таблице системой STATISTICA.

Таблица 2.6

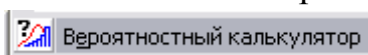
Корреляционная матрица ассоциации микроэлементов в почво-грунтах по методу ISP.

	Na _I	Fe _I	Ba _I	Sc _I	Cr _I	Co _I	La _I
Na _I	1,00	0,50	-0,22	0,74	0,00	0,24	0,47
Fe _I	0,50	1,00	0,12	0,82	0,13	0,46	0,67
Ba _I	-0,22	0,12	1,00	-0,07	0,53	0,55	0,07
Sc _I	0,74	0,82	-0,07	1,00	-0,02	0,31	0,72
Cr _I	0,00	0,13	0,53	-0,02	1,00	0,60	-0,13
Co _I	0,24	0,46	0,55	0,31	0,60	1,00	0,42
La _I	0,47	0,67	0,07	0,72	-0,13	0,42	1,00


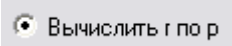
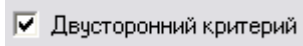
В тоже время в системе STATISTICA предоставлена возможность вычисления критического коэффициента корреляции Пирсона по заданному уровню значимости и объему выборки. Для этого необходимо:

Шаг 1. Запустите в главном меню модуль  .

Шаг 2. В стартовой панели модуля нажмите клавишу



Шаг 3. В появившемся окне выберите  (рис.2.71)

Шаг 4. В окне  введите значения объема выборки $N = 91$ и уровень значимости $p = 0,05$, выделите точкой  и галочкой  и нажмите

кнопку  (рис.2.72). В окне  появится значение $r =$

0,206098 - критическое значение – граница значимых, то есть отличных

от ноля на уровне значимости 0,05 при объеме выборки $N = 91$, коэффициентов парной корреляции Пирсона.

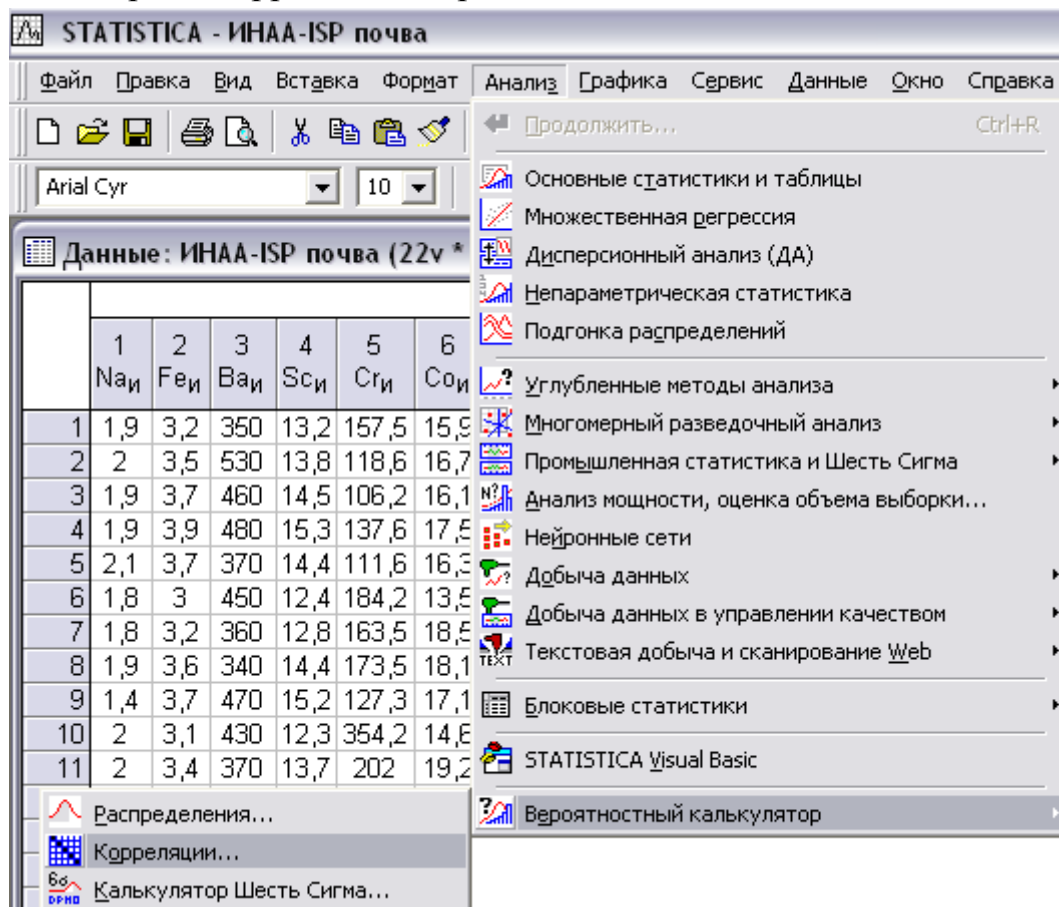


Рис. 2.71. Выбор Вероятностный калькулятор (Корреляции...) в модуле Анализ

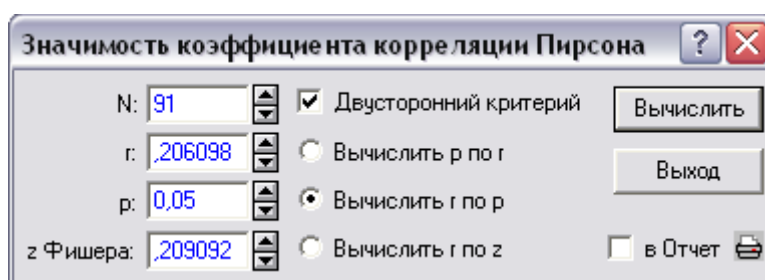

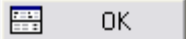


Рис. 2.72. Вычисление коэффициента корреляции по объему выборки при заданном уровне значимости.

Оценим значимость различия между двумя значимыми коэффициентами корреляции одной и той же пары микроэлементов (например, Fe и Sc в разных методах: 0,87 в ИНАА и 0,82 в ISP):

Шаг 1. Запустите в главном меню модуль Анализ.

Шаг 2. Далее выбрать Основные статистики и таблицы.

Шаг 3. В появившемся окне выберите  (рис.2.73) и нажмите .

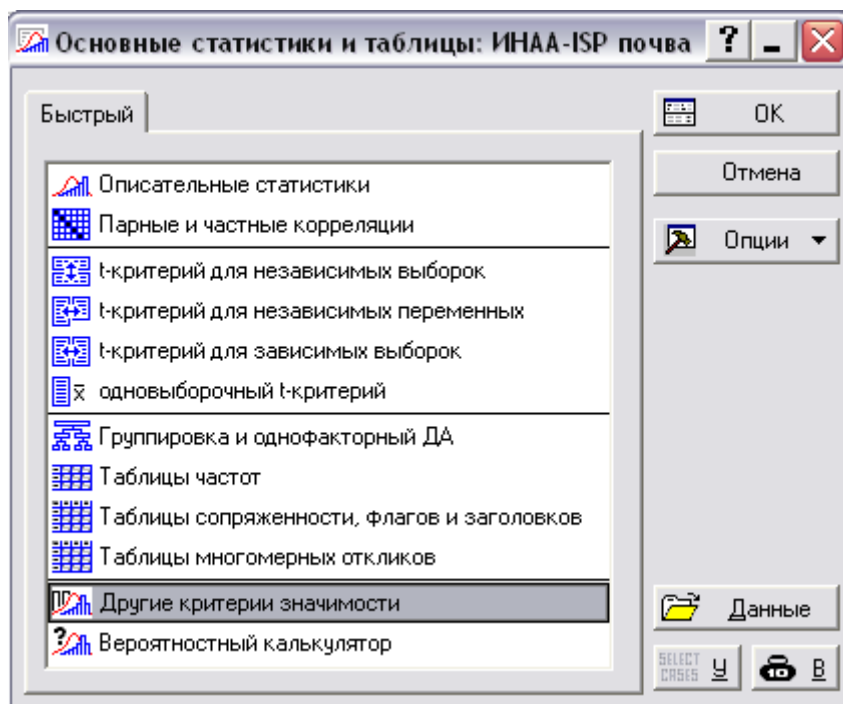



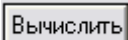


Рис. 2.73. Выбор  в модуле .

Шаг 4. В окне  введите значения сравниваемых значений коэффициентов корреляции (напр., $r_1 = 0,87$ и $r_2 = 0,82$), объемов выборок ($N_1 = 91$ и $N_2 = 91$), выделите точкой двухсторонний режим и нажмите кнопку . Здесь же в окне появится значение $p = 0,2439$ (рис.2.74).

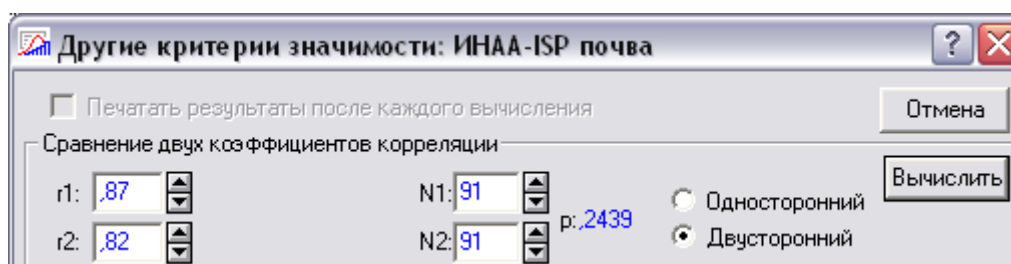


Рис. 2.74. Оценка расхождения между двумя коэффициентами корреляции Пирсона

То есть данные $r_1 = 0,87$ и $r_2 = 0,82$ при $N_1 = 91$ и $N_2 = 91$ различаются не значимо ($p \approx 0,24 > 0,10$).




2.4.2. Вычисление коэффициента корреляции Спирмена

Корреляция между результатами содержаний двух химических элементов может характеризоваться прежде всего коэффициентом парной корреляции Пирсона r , отражающим степень линейной связи, для нормальных распределений или ранговым коэффициентом корреляции Спирмена R , если распределения заметно отличается от нормального. При этом оценка значимости корреляции приводит к уровню значимости для коэффициента корреляции, определяющему значимость отличия коэффициента корреляции от нуля (значимость корреляции). Представляет интерес возможность определения коэффициента корреляции Спирмена R средствами пакета STATISTICA 6.1. Для этого необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль .

Шаг 2. В стартовой панели модуля нажмите клавишу

  (рис.2.2, 2.34, 2.71).

Шаг 3. В появившемся окне  выберите  (рис.2.75) и нажмите .

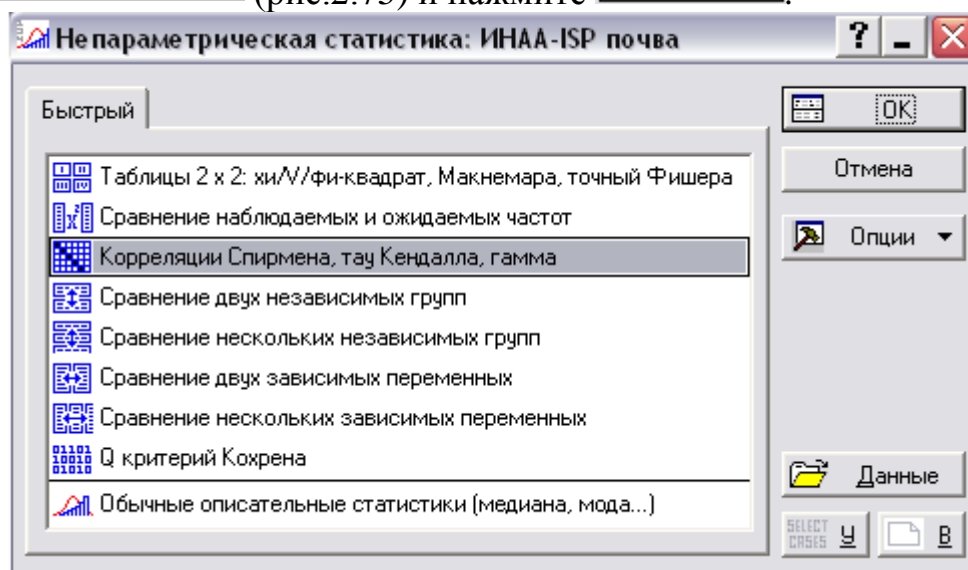


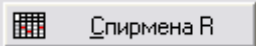


Рис. 2.75. Выбор  в модуле



Шаг 4. В открывшемся окне нажмите кнопку  и выберите 4-е переменные базы данных, например, Fe_{II} , Sc_{II} , Fe_{III} и Sc_{III} (рис.2.76) и нажмите .

Шаг 5. После этого нажмите кнопку  (рис.2.76) для просмотра результатов вычисления парных коэффициентов корреляции Спирмена (рис.2.77).

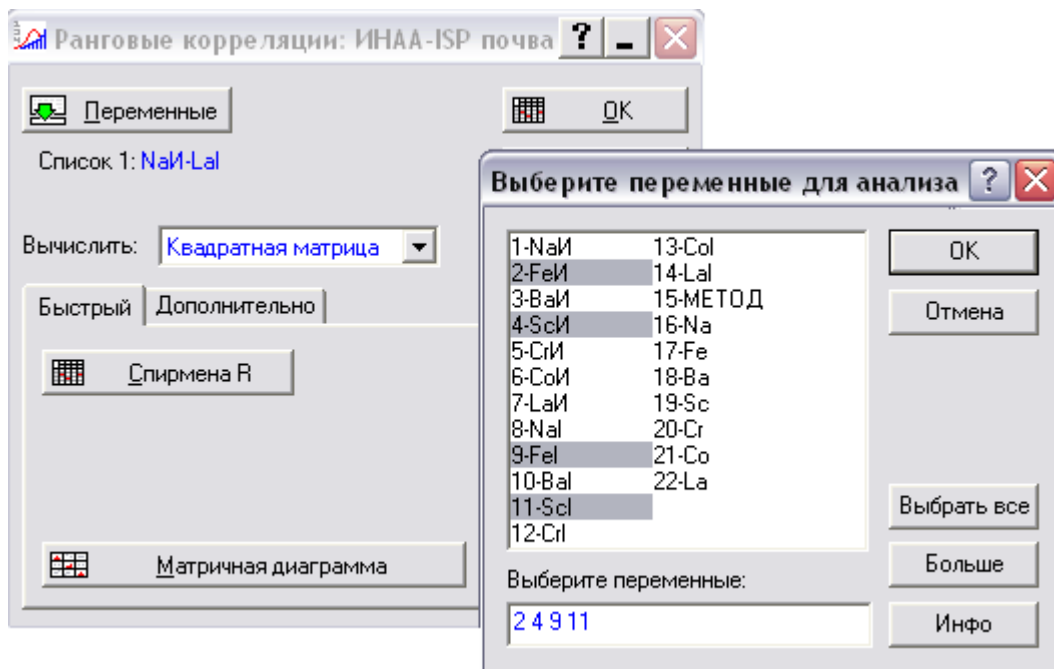


Рис. 2.76. Оформление вычисления R Спирмена

Ранговые корреляции Спирмена (ИНАА-ISP почва)				
ПД попарно удалены				
Отмеченные корреляции значимы на уровне $p < ,05000$				
Перем.	Fe _И	Sc _И	Fe _І	Sc _І
Fe _И	1,000000	0,81053	0,74545	0,53112
Sc _И	0,81053	1,000000	0,60546	0,70594
Fe _І	0,74545	0,60546	1,000000	0,70654
Sc _І	0,53112	0,70594	0,70654	1,000000

Рис. 2.77. Результаты вычисления R Спирмена.

Оценка значимости различия между двумя коэффициентами корреляции Спирмена пары Fe и Sc в разных методах ($\approx 0,81$ в ИНАА и $\approx 0,71$ в ISP) по выше приведенной схеме (рис.2.74) приводят к выводу о не значимом ($p \approx 0,11$) различии в $R_{И} \approx 0,81$ и $R_{І} \approx 0,71$.

Оценка степени различия коэффициентов корреляции R Спирмена и r Пирсона пары микроэлементов Fe и Sc в рамках одного метода приводят к выводу о не значимом ($p \approx 0,17$) различии в $r_{И} \approx 0,87$ и $R_{И} \approx 0,81$ и слабо значимом ($p \approx 0,075$) различии в $r_{І} \approx 0,82$ и $R_{І} \approx 0,71$.

2.4.3. Построение диаграммы рассеяния и линии регрессии

Во второй части **этапа III** для построения составных диаграмм рассеяния и уравнений регрессий для наиболее значимых корреляционных связей в ассоциациях необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню (полоса меню) модуль **Графика**.

Шаг 2. Выберите **Категоризованные графики** и далее щелкните мышкой на разновидности **Диаграммы рассеяния...** (рис.2.78).

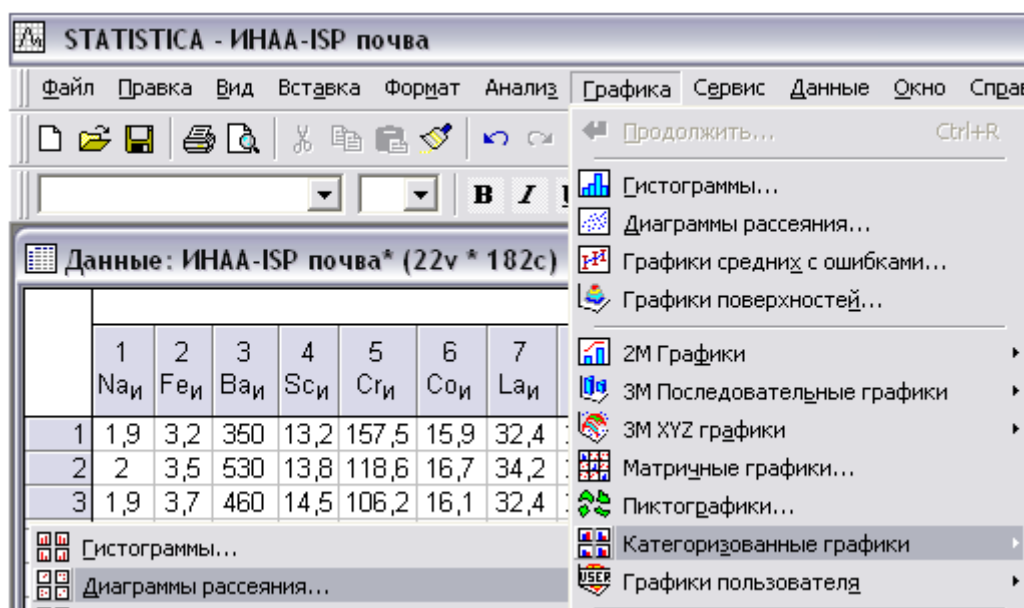


Рис. 2.78. Выбор **Диаграммы рассеяния...** в модуле **Графика**.

Шаг 3. В появившемся окне щелкните **Дополнительно**.

Шаг 4. В открывшемся окне нажмите кнопку **Переменные:** и выберите переменные категориальной базы данных (например, Fe – в окне **X:**, Sc – в окне **Y:** и переменная Метод – в окне **X-Категория:**) и нажмите **OK**.

Шаг 5. Далее в окне **X-Категории** выделите режим **Коды:**, нажмите кнопку **Задать коды**, выберите **Все** в **Коды категорий:** и **OK**.

Шаг 6. После этого в окне **Размещение** выделите режим **Вместе**, в окне **Статистики** выделите режим **Корреляция и r-уров.**, в окне **Тип подгонки:** выделите режим **Линейная** (рис.2.79) и нажмите кнопку **OK**.

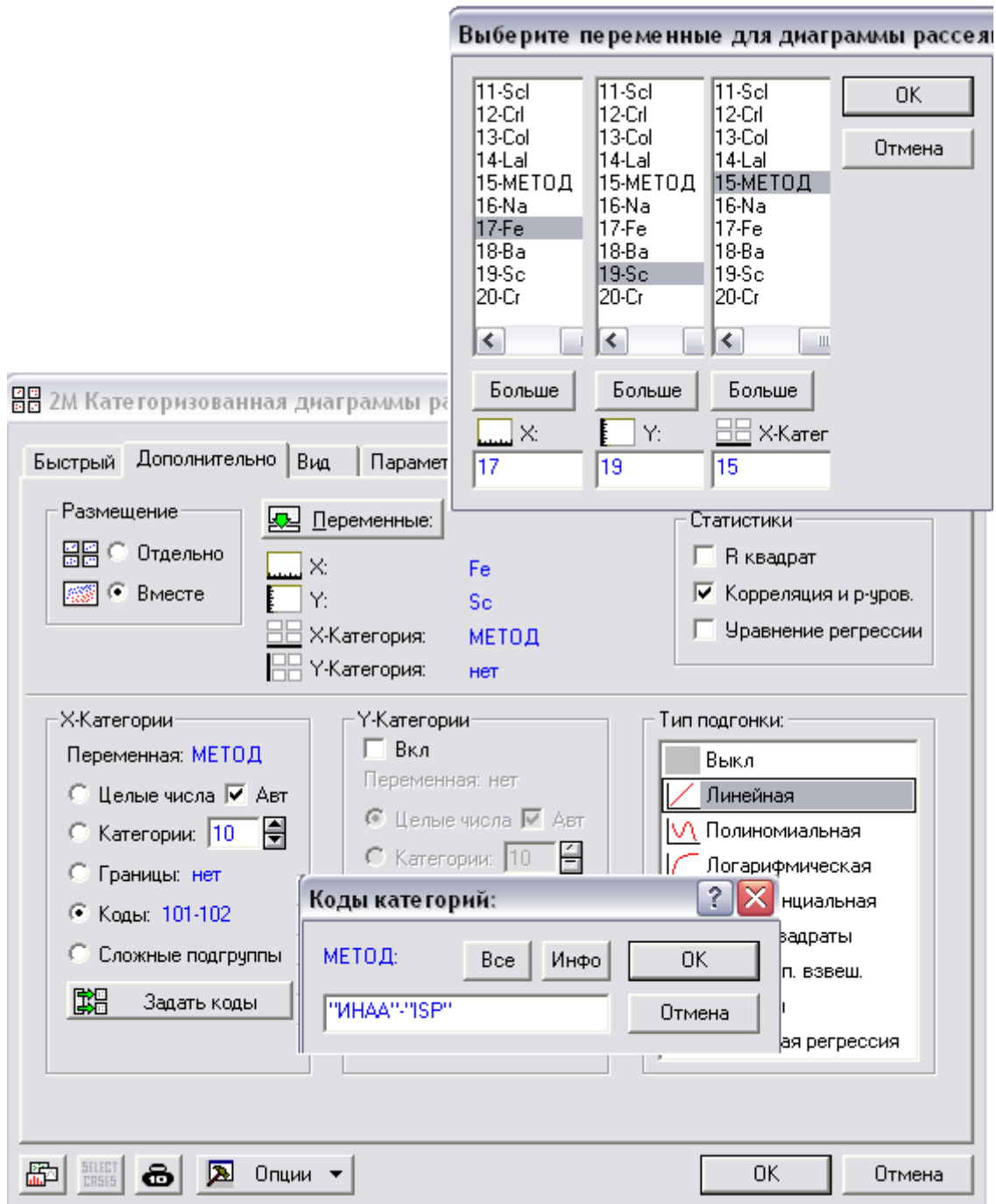


Рис. 2.79. Окно **2M Категоризованная диаграммы рассеяния**

Составная диаграмма рассеяния с прямыми линиями составных регрессий Fe : Sc по методам ИНАА и ISP появится на экране (рис.2.80). Попутно приводятся коэффициенты корреляции Пирсона с указанием их уровней значимости и уравнения регрессии зависимости Fe и Sc по методам ИНАА и ISP. В данном случае к прямой регрессии можно до-

бавить 95% эллипс рассеяния и границы 95% доверительного интервала. Для этого достаточно щелкнуть дважды мышкой в поле диаграммы, в появившемся окне **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**.

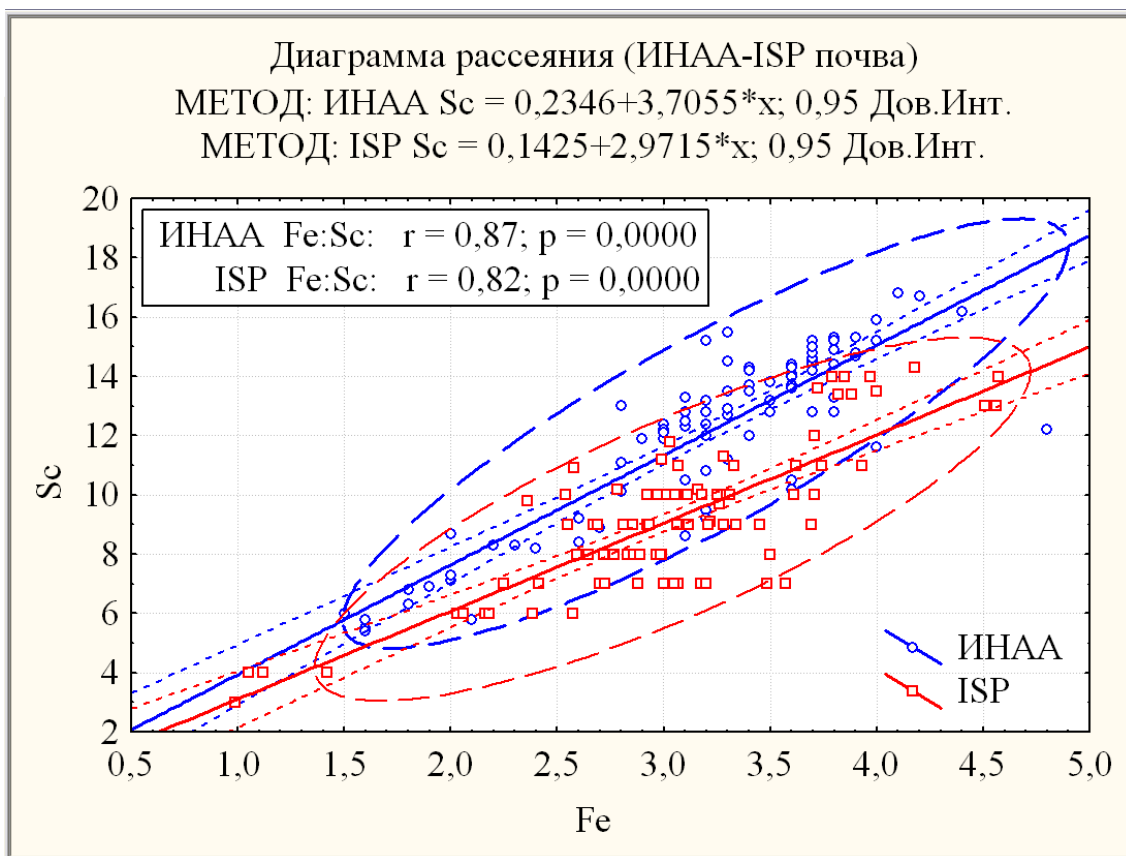


Рис. 2.80. Составная диаграмма рассеяния зависимости Fe и Sc по методам ИНАА и ISP

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. На диаграммах рассеяния пунктирной замкнутой линией нарисован 95% эллипс рассеяния, а мелко пунктирной линией – 95% доверительный интервал линии регрессии.

Если в окне **Размещение** выделить режим **Отдельно** (Шаг.6), то на экране появятся отдельные диаграммы зависимости Fe и Sc по методам ИНАА и ISP (рис.2.81).

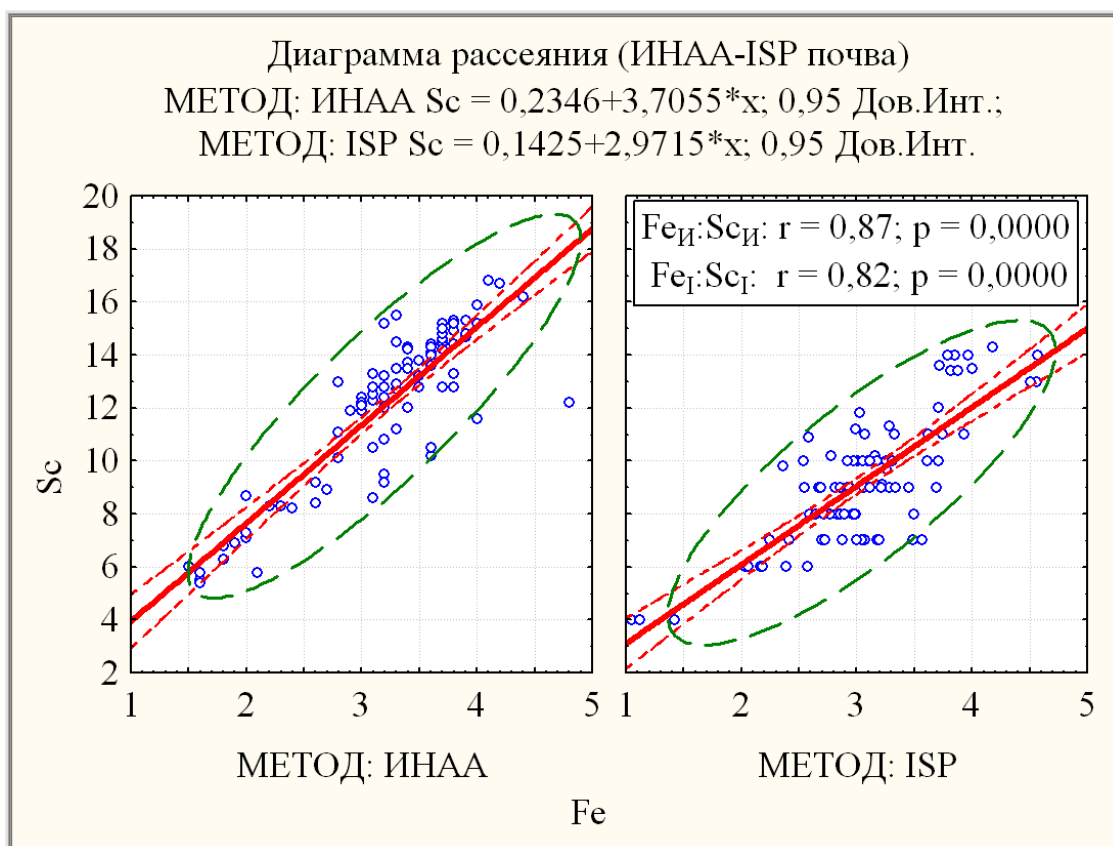


Рис. 2.81. Раздельные диаграммы рассеяния зависимости Fe и Sc по методам ИНАА и ISP

Средствами пакета STATISTICA возможно построение диаграммы рассеяния с гистограммами. Для этого необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль **Графика**.

Шаг 2. В стартовой панели модуля выберите **2М Графики** и щелкните мышкой на **Диаграммы рассеяния с гистограммами...** (рис.2.82).

Шаг 3. В появившемся окне щелкните **Дополнительно**.

Шаг 4. В открывшемся окне нажмите кнопку **Переменные:** и выберите переменные раздельной базы данных, например, переменная $Fe_{И}$ – в окне **X:** и переменная $Sc_{И}$ – в окне **Y:** (рис.2.83) и нажмите **ОК**.

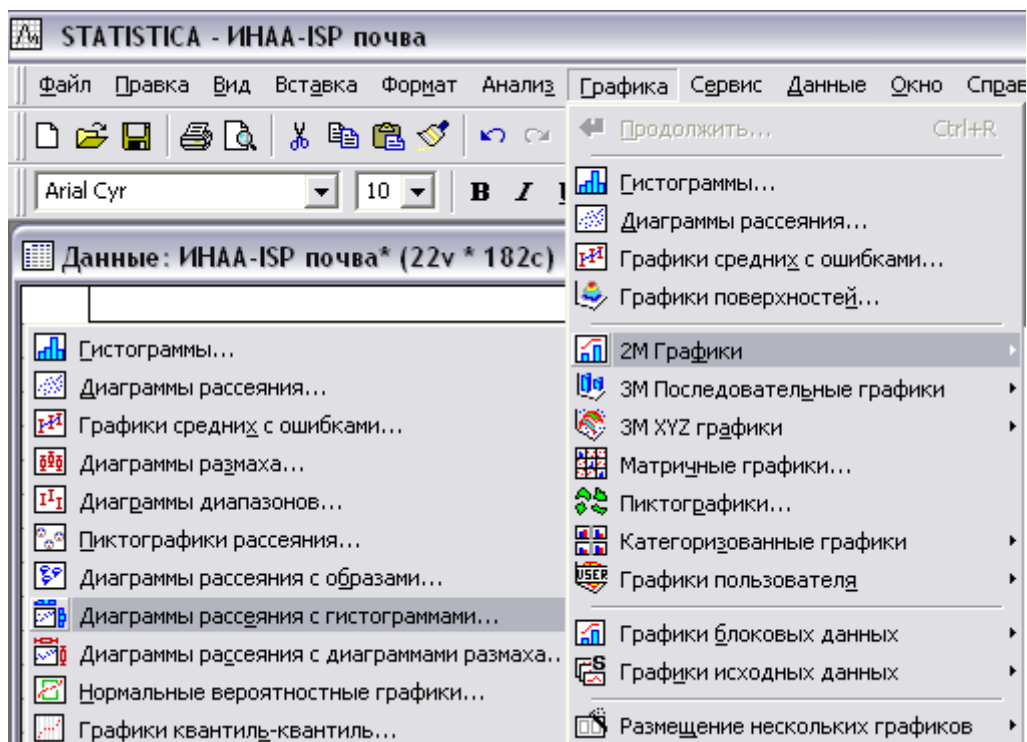




Рис. 2.82. Выбор  Диаграммы рассеяния с гистограммами... в модуле  Графика

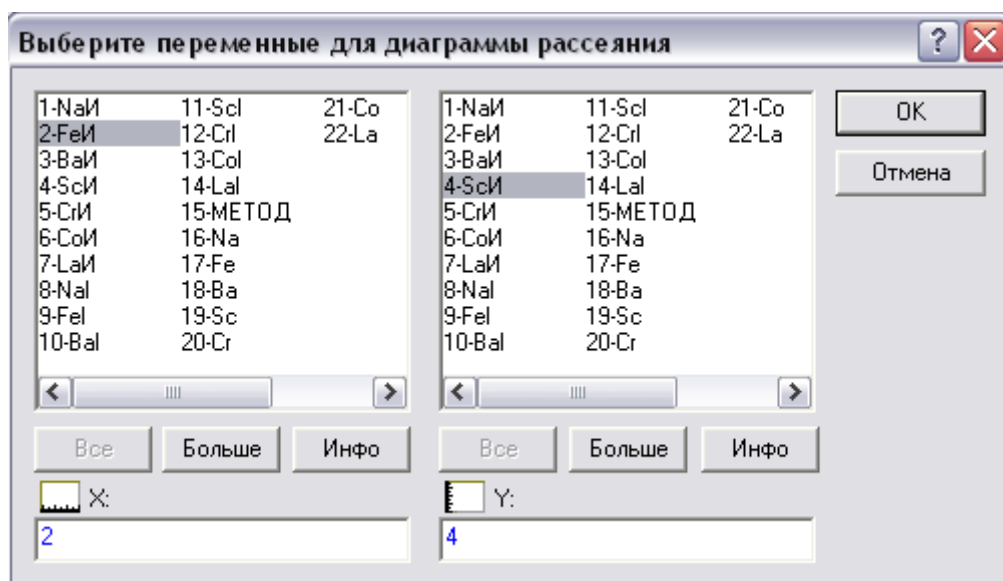
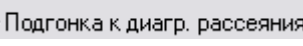

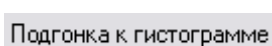

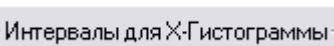
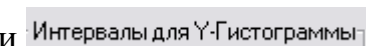
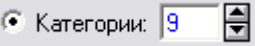


Рис. 2.83. Выбор  Переменные:

Шаг 5. После этого в окне  Подгонка к диагр. рассеяния выделите режим  Линейная, в окне  Подгонка к гистограмме выделите  Нормальное, в окне  Интервалы для X-Гистограммы и  Интервалы для Y-Гистограммы выделите  Категории:

в окне **Статистики гистограммы** выделите **Критерий Шапиро-Уилка** и **Критерий К-С**, в окне **Статистики диаг. рассеяния** — **Корреляция и р-уровень** и **Уравнение регрессии** (рис.2.84) и нажмите кнопку **OK**.

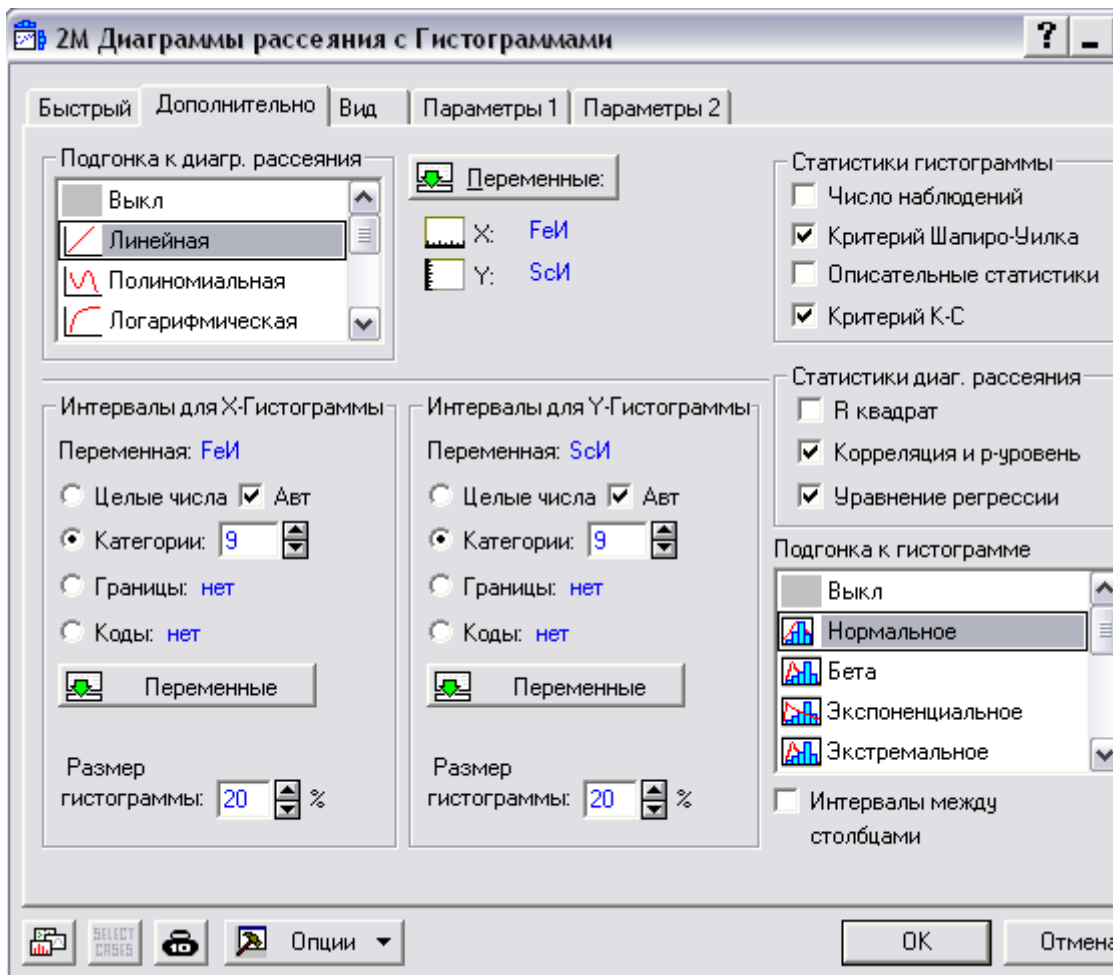


Рис. 2.84. Окно **2M Диаграммы рассеяния с Гистограммами**

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 **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 **OK**.

Следующая диаграмма рассеяния с гистограммами появится на экране (рис.2.85). Аналогично для метода ISP (рис.2.86):

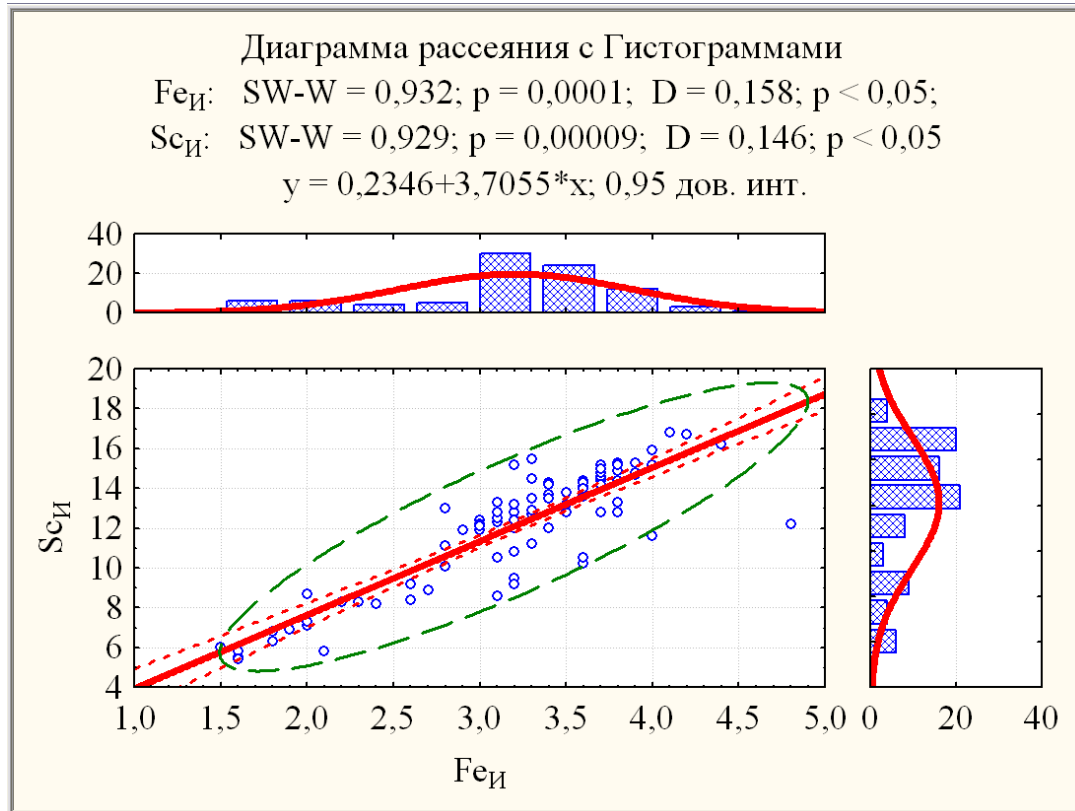


Рис. 2.85. Диаграмма рассеяния зависимости Fe_{II} и Sc_{II} с гистограммами

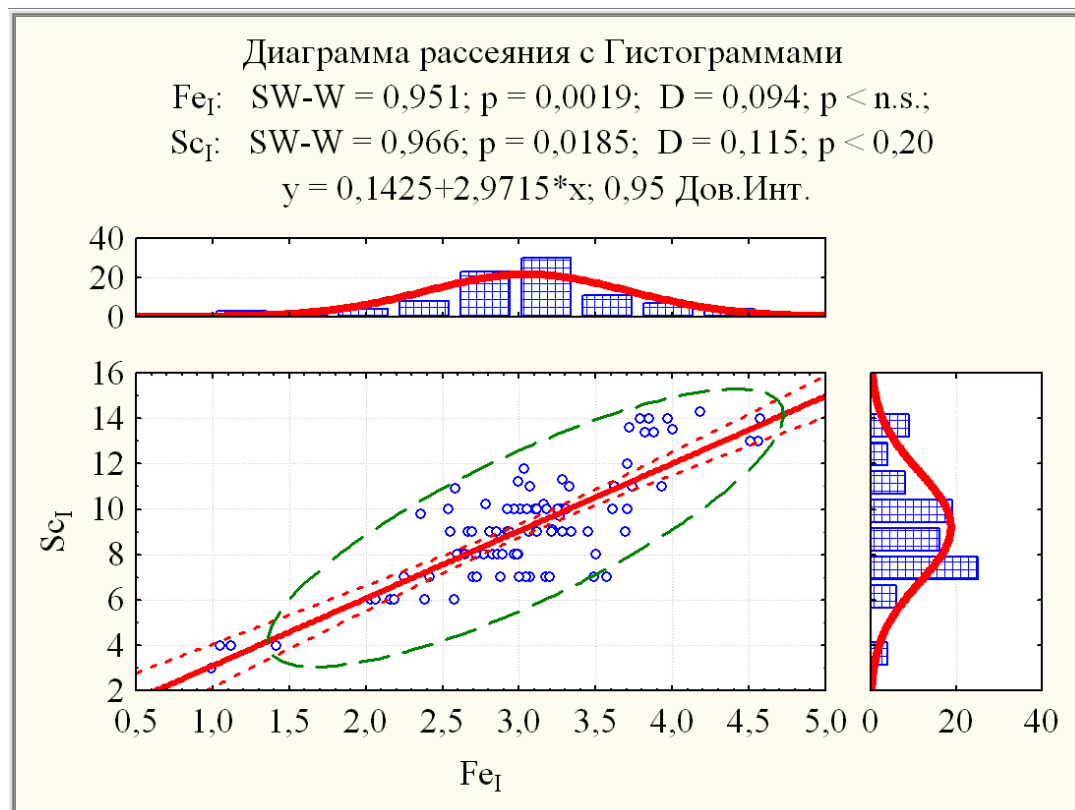


Рис. 2.86. Диаграмма рассеяния зависимости Fe_I и Sc_I с гистограммами

Диаграммы рассеяния позволяют также находить «выбросы» (нетипичные аномальные данные), которые вносят свой вклад в закон распределения, в значение коэффициента корреляции Пирсона и в уравнение линейной регрессии. Даже один выброс может значительно изменить коэффициент корреляции Пирсона. Наглядным примером тому является распределение содержания C_T с выбросом (рис.2.87) с координатами (300; 350), что порождает статистически значимый ($p = 0,02$) коэффициент корреляции Пирсона $r = 0,25$, и без него (рис.2.88), что приводит к не значимому ($p = 0,43$) коэффициенту корреляции Пирсона $r = 0,08$. При этом заметно меняется также уравнение линейной регрессии и положение прямой регрессии.

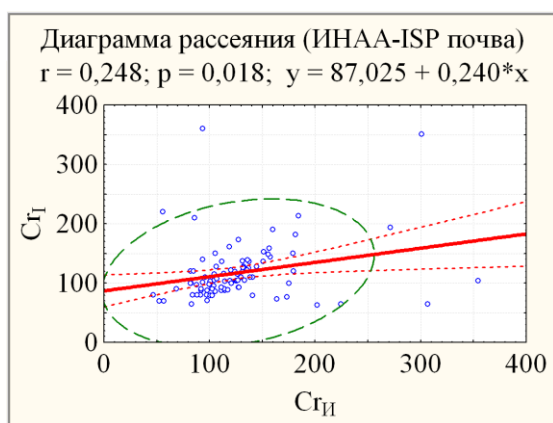


Рис. 2.87. Диаграмма рассеяния C_T с выбросом в т. (300; 350).

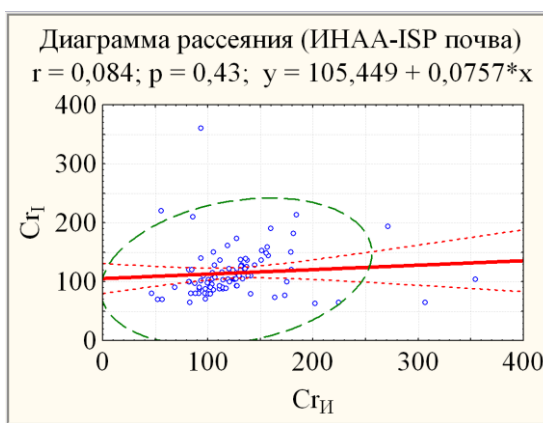



Рис. 2.88. Диаграмма рассеяния C_T без выброса.

В системе STATISTICA существует средство  – **Закрашивание** среди набора **Графические инструменты**, позволяющее интерактивно удалять выбросы и непосредственно наблюдать за изменением значения коэффициента корреляции Пирсона, уравнения линейной регрессии и положения линии регрессии (рис.2.89).

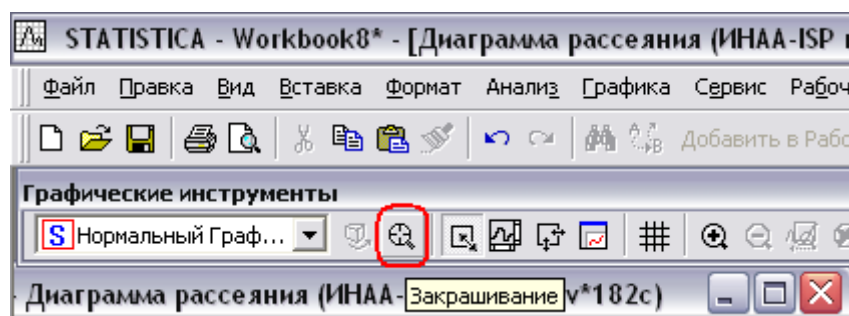





Рис. 2.89. Активация средства  – **Закрашивание** для удаления выбросов

Для активации  необходимо после появления на экране диаграммы рассеяния распределения содержания C_T с выбросом (рис. 2.87)

нажать в головной панели кнопку - ярлык  в форме лупы (рис.2.89). Затем появившемся курсором в форме лупы навести на выброс с координатами (300; 350) и щелкнуть мышкой, в результате чего появится диаграмма рассеяния распределения содержания Cr без выброса (рис. 2.88).

Наконец, диаграммы рассеяния позволяют понять, действительно ли между двумя исследуемыми переменными X и Y имеется связь, особенно, если она есть и является нелинейной. Проиллюстрируем построение уравнения нелинейной регрессии и оценку степени адекватности регрессионной модели (см. конец п.1.2.4) на примере распределение содержания Ва (рис. 2.90).

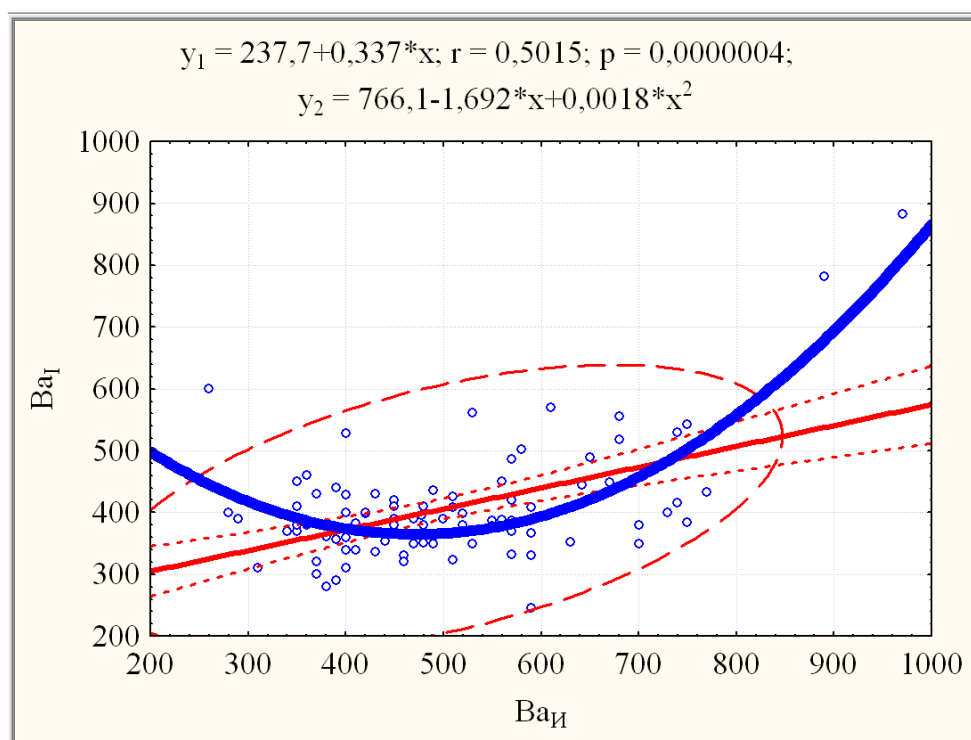



Рис. 2.90. Кривые линейной и квадратичной регрессий Ва

Кривую квадратичной регрессии и ее уравнение можно построить отдельно аналогично прямой регрессии, заказав в окне **Тип подгонки:** (рис. 2.79) тип  **Полиномиальная**. Для совмещения линий регрессий сначала можно построить диаграмму рассеяния с прямой регрессии (аналогично рис. 2.87), а затем добавить кривую квадратичной регрессии с помощью **Все параметры графика**, нажав последовательно **Пользовательская функция**, затем **Новая функция** и указав в **Браузер функций** уравнение квадратичной регрессии.

Для оценки значимости регрессии необходимо выполнить последовательность следующих действий:

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

Шаг 2. Выберите **Углубленные методы анализа** в стартовой панели модуля и далее щелкните **Общие регрессионные модели** (рис.2.91)

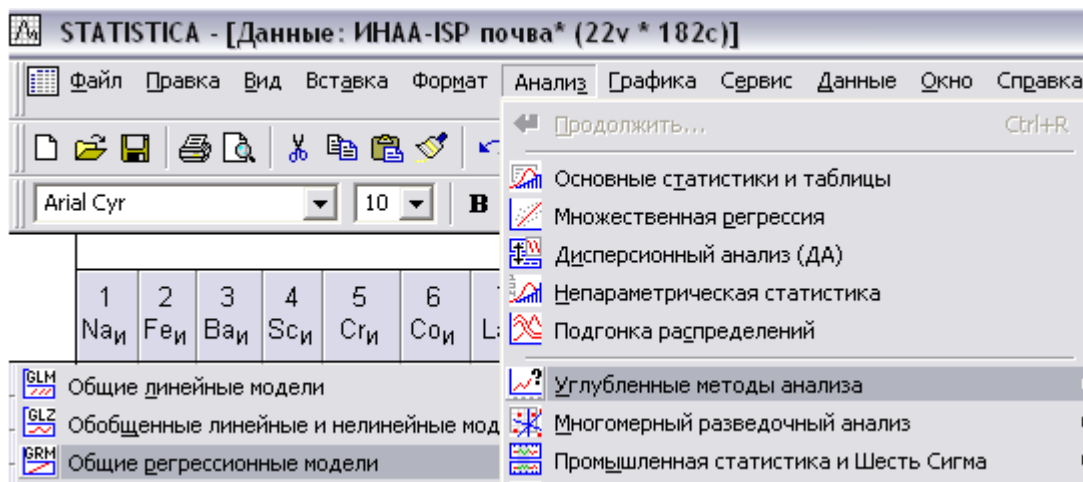


Рис. 2.91. Выбор **Общие регрессионные модели** в модуле **Анализ**

Шаг 3. В окне **Вид анализа:** щелкните мышкой на разновидности **Полиномиальная регрессия** (рис.2.92).

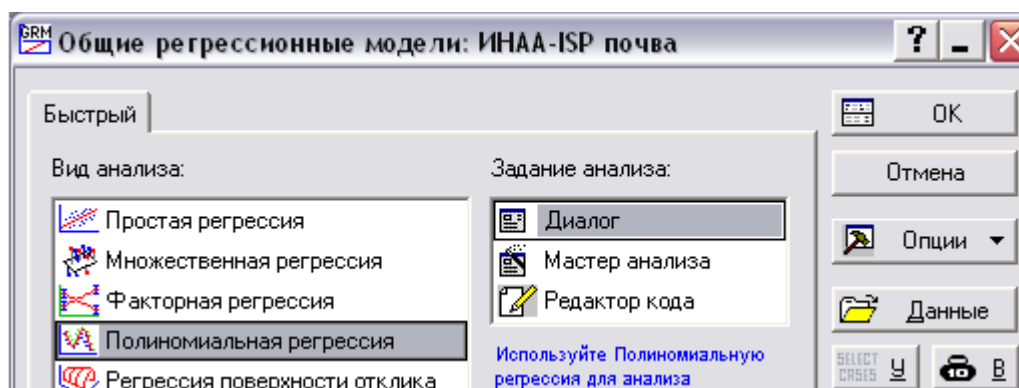


Рис. 2.92. Выбор **Полиномиальная регрессия** в модуле

Шаг 4. В открывшемся окне нажмите кнопку **Переменные** и выберите переменные «Ba_и» в левом окне и «Ba_г» в правом окне и нажмите ОК. Далее в окне **Межгруп. эффекты** закажите порядок полинома (рис.2.93) нажмите **ОК**.

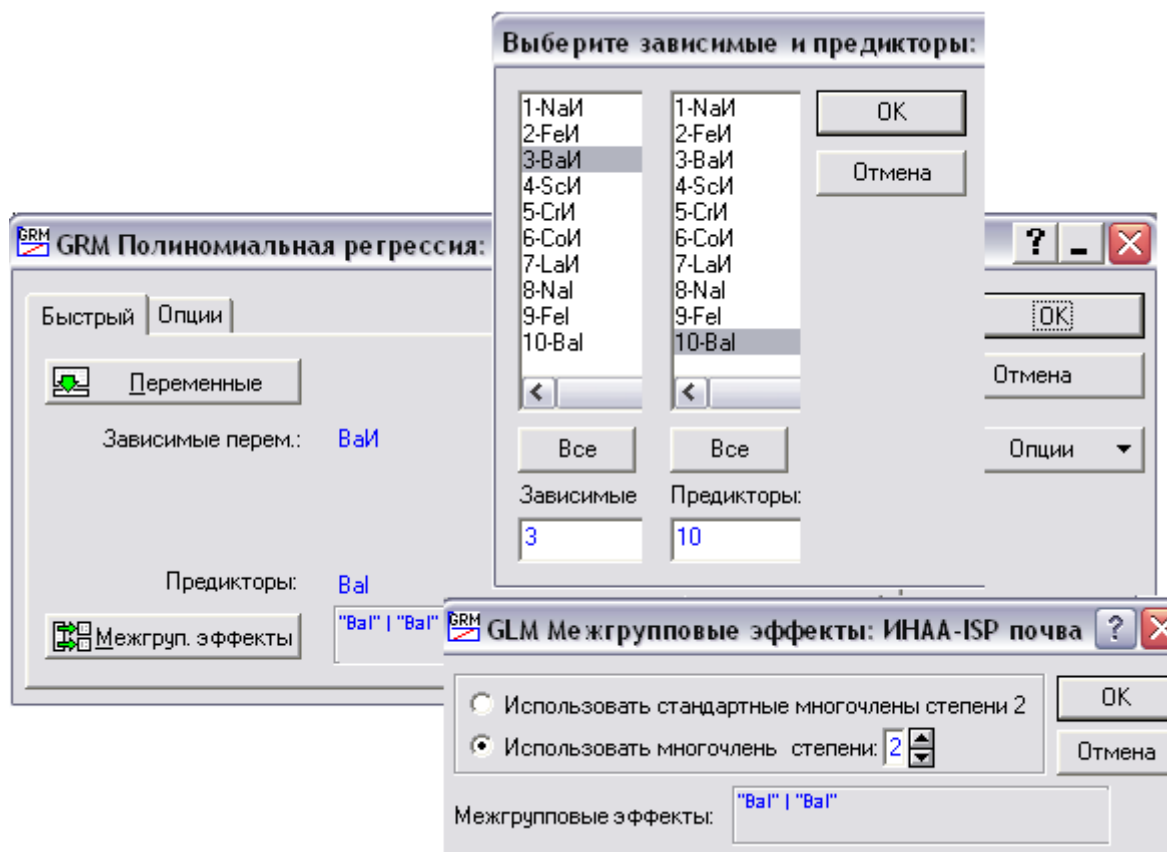




Рис. 2.93. Выбор  Переменные и порядка полинома.

Шаг 5. Нажмите кнопку  Все эффекты в открывшемся окне

 GRM Результаты (рис.2.94).

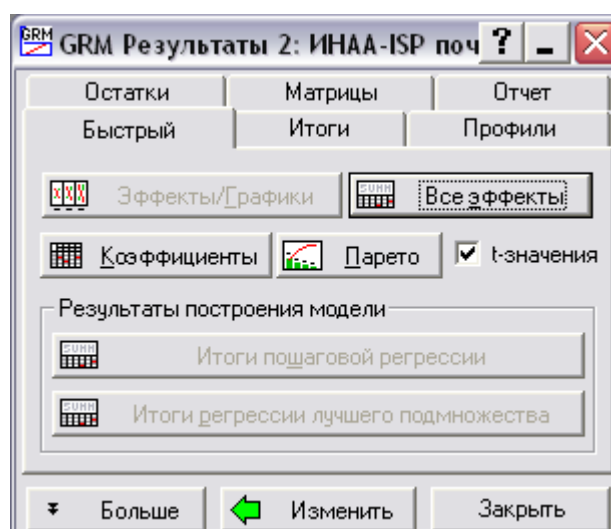


Рис. 2.94. Окно просмотра  GRM Результаты

Следующая таблица появится на экране:

Таблица 2.7

Оценки значимости квадратичной регрессионной зависимости V_{aII} и V_{aISP} (F-критерий и коэффициента детерминации).

	R^2	\hat{R}^2	SS модель	m	MS модель	SS остаток	сс ост	MS остаток	F	p_F
V_{aII}	0,273	0,256	463164,0	2	231582,0	1235774	88	14042,89	16,49	0,000001

Наиболее интересны значения скорректированного коэффициента детерминации $\hat{R}^2 = 0,256$, порядка полинома $m = 2$, значение $F = 16,49$, и уровень его значимости $p_F = 0,000001$. Скорректированный коэффициент детерминации \hat{R}^2 измеряет качество построенной регрессии: чем он ближе к 1, тем лучше регрессия «объясняет» зависимость в данных. Оценка значимости регрессии (качество уравнения регрессии) проверяется с помощью F -критерия

$$F = \frac{R^2}{1-R^2} \frac{N-M-1}{M}, = \frac{0,273}{1-0,273} \cdot \frac{91-2-1}{2} \approx 16,49.$$

С другой стороны, в рамках дисперсионного анализа мы имеем разложение общей вариации на составляющие, обусловленные регрессией ($SS_{\text{мод}}$) и ошибкой ($SS_{\text{ост}}$), отношение средних квадратов отклонений которых также определяет значение F -критерия: $F = MS_{\text{мод}} / MS_{\text{ост}} = 231582,0 / 14042,89 \approx 16,49$.

Перебирая m , получим таблицу 2.8.

Таблица 2.8

Оценки значимости полиномиальной регрессионной зависимости V_{aII} и V_{aISP} (F-критерий и коэффициента детерминации).

m	R^2	\hat{R}^2	$F(R^2)$	p_F
1	0,25147	0,24306	29,89988	0,000000
2	0,27262	0,25609	16,49105	0,000001
3	0,27281	0,24774	10,87957	0,000004


В табл.2.8 приведена зависимость коэффициента детерминации, значения F-критерия и уровня его значимости p_F от порядка m полинома. Как следует из табл.2.8, криволинейные регрессии несущественно улучшают линейную регрессию (\hat{R}^2 далек от 1, а $p_F < 0,001$), то есть и криволинейные регрессии высоко значимо отличаются от экспериментального распределения зависимости V_{aII} и V_{aI} .



2.5. Канонический анализ двух ассоциаций микроэлементов

В рамках **IV этапа** многомерного статистического анализа содержаний микроэлементов, полученных двумя методами, проведем канонический анализ соответствующих двух ассоциаций микроэлементов.

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, полученных методом ИНАА и ISP, в почве территории юго-восточного сектора влияния Сибирского химического комбината (измерения на территории г. Томска №20–80в приложении П₁). Соответствующий фрагмент выделен из П₁ в файл с названием ИНАА-ISP почва Томск.sta

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

Шаг 1. Запустите в полосе меню головной модуль  Анализ и в нем модуль  Канонический анализ (см. рис. 2.95).

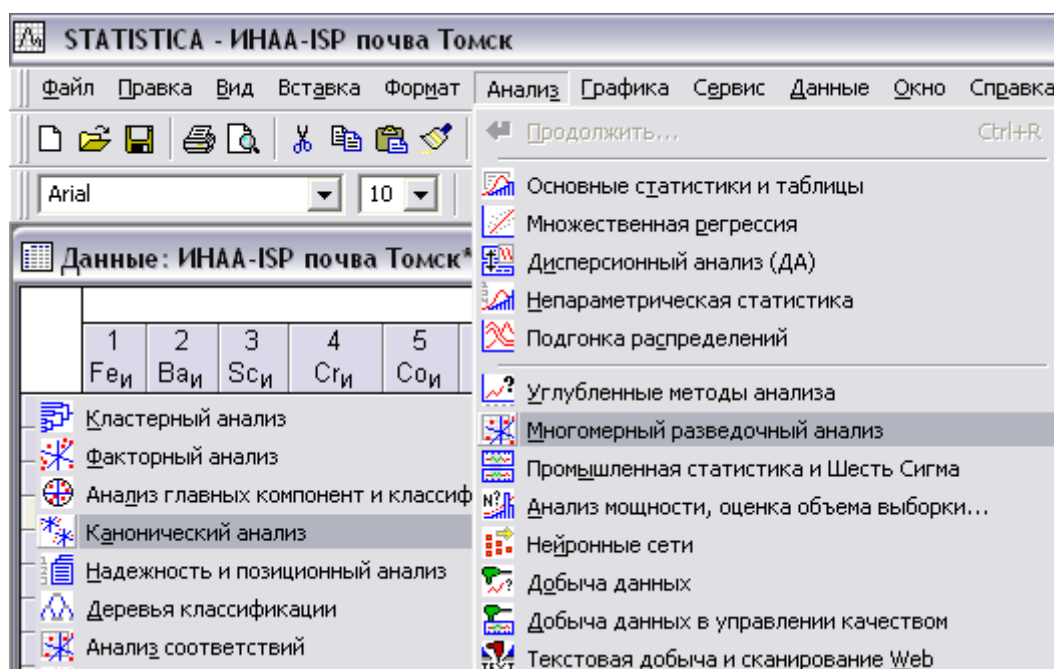





Рис. 2.95. Модуль  Канонический анализ в модуле  Анализ

Шаг 2. Выберите базы данных, как показано на рис. 2.96, и нажмите  в каждом окне.

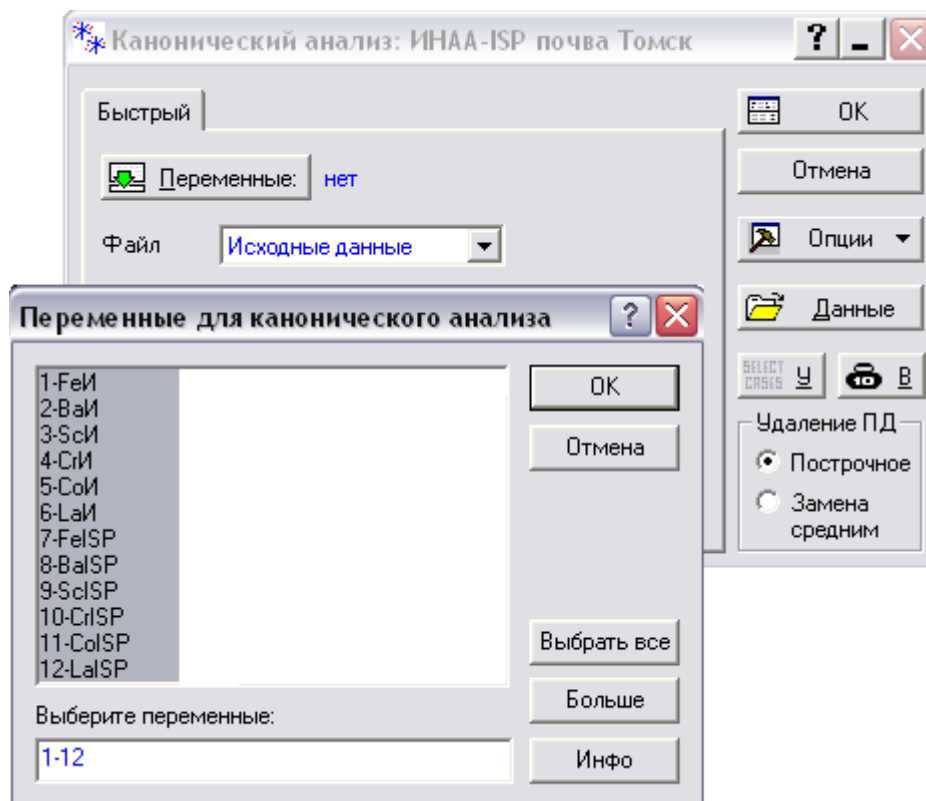


Рис. 2.96. Выбор исследуемых переменных для канонического анализа

На экране появится окно **Определение модели** (рис. 2.97).

Шаг 3. В окне **Определение модели** (рис. 2.97) можно просмотреть результаты вычисления выборочных числовых характеристик:

- ◇ **Средние и стандартные отклонения** - векторов математического ожидания и среднеквадратичного отклонения (рис. 2.98);
- ◇ **Корреляционная матрица** - матрицы парных коэффициентов корреляции (рис. 2.99);

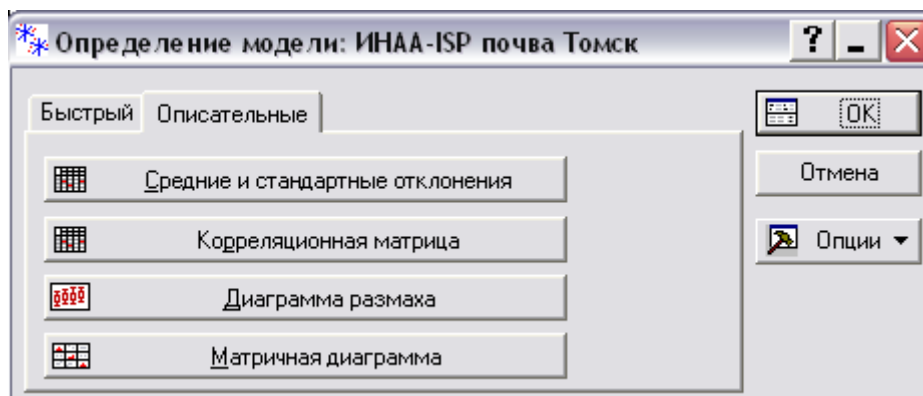




Рис. 2.97. Режим **Описательные** в окне **Определение модели**



	Средние и станд. отклонения	
	Средние	Ст. Откл.
Fe _И	3,1344	0,6928
Ba _И	541,3115	140,1394
Sc _И	11,4279	2,7356
Cr _И	120,1148	39,6226
Co _И	15,6721	8,2371
La _И	25,8557	4,5585
Fe _{ISP}	3,0450	0,5748
Ba _{ISP}	425,7869	104,7652
Sc _{ISP}	8,5574	1,8932
Cr _{ISP}	121,1803	54,5015
Co _{ISP}	19,3934	9,5748
La _{ISP}	21,9016	5,9964


Рис. 2.98. Векторы выборочных математического ожидания и средне-квadraticного отклонения

Корреляции (ИНАА-ISP почва Томск)												
	Fe _И	Ba _И	Sc _И	Cr _И	Co _И	La _И	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Fe _И	1,000	0,216	0,845	0,174	0,302	0,552	0,559	0,029	0,313	-0,308	0,221	0,635
Ba _И	0,216	1,000	0,058	0,360	0,418	-0,097	0,034	0,469	-0,115	0,189	0,344	0,112
Sc _И	0,845	0,058	1,000	-0,059	0,130	0,698	0,491	-0,236	0,480	-0,407	0,141	0,747
Cr _И	0,174	0,360	-0,059	1,000	0,823	-0,167	-0,069	0,550	-0,256	0,481	0,500	-0,087
Co _И	0,302	0,418	0,130	0,823	1,000	0,013	0,152	0,579	0,010	0,429	0,685	0,130
La _И	0,552	-0,097	0,698	-0,167	0,013	1,000	0,484	-0,172	0,492	-0,209	0,097	0,601
Fe _{ISP}	0,559	0,034	0,491	-0,069	0,152	0,484	1,000	0,147	0,824	0,037	0,331	0,566
Ba _{ISP}	0,029	0,469	-0,236	0,550	0,579	-0,172	0,147	1,000	-0,039	0,579	0,579	0,005
Sc _{ISP}	0,313	-0,115	0,480	-0,256	0,010	0,492	0,824	-0,039	1,000	-0,080	0,195	0,670
Cr _{ISP}	-0,308	0,189	-0,407	0,481	0,429	-0,209	0,037	0,579	-0,080	1,000	0,676	-0,338
Co _{ISP}	0,221	0,344	0,141	0,500	0,685	0,097	0,331	0,579	0,195	0,676	1,000	0,204
La _{ISP}	0,635	0,112	0,747	-0,087	0,130	0,601	0,566	0,005	0,670	-0,338	0,204	1,000

Рис. 2.99. Матрица выборочных парных коэффициентов корреляции

В рамках визуализации корреляционной матрицы исследуемых переменных можно построить диаграмму размахов  Диаграмма размаха и матрицу диаграмм рассеяния с гистограммами  Матричная диаграмма (рис. 2.100):

Шаг 4. В окне  Определение модели (рис. 2.97) нажмите кнопку  Матричная диаграмма, в открывшемся окне выберите некоторые перемен-

ные, например, Fe_{II} , Sc_{II} , Cr_{ISP} и Co_{ISP} (аналогично рис. 2.96) и нажмите кнопку .

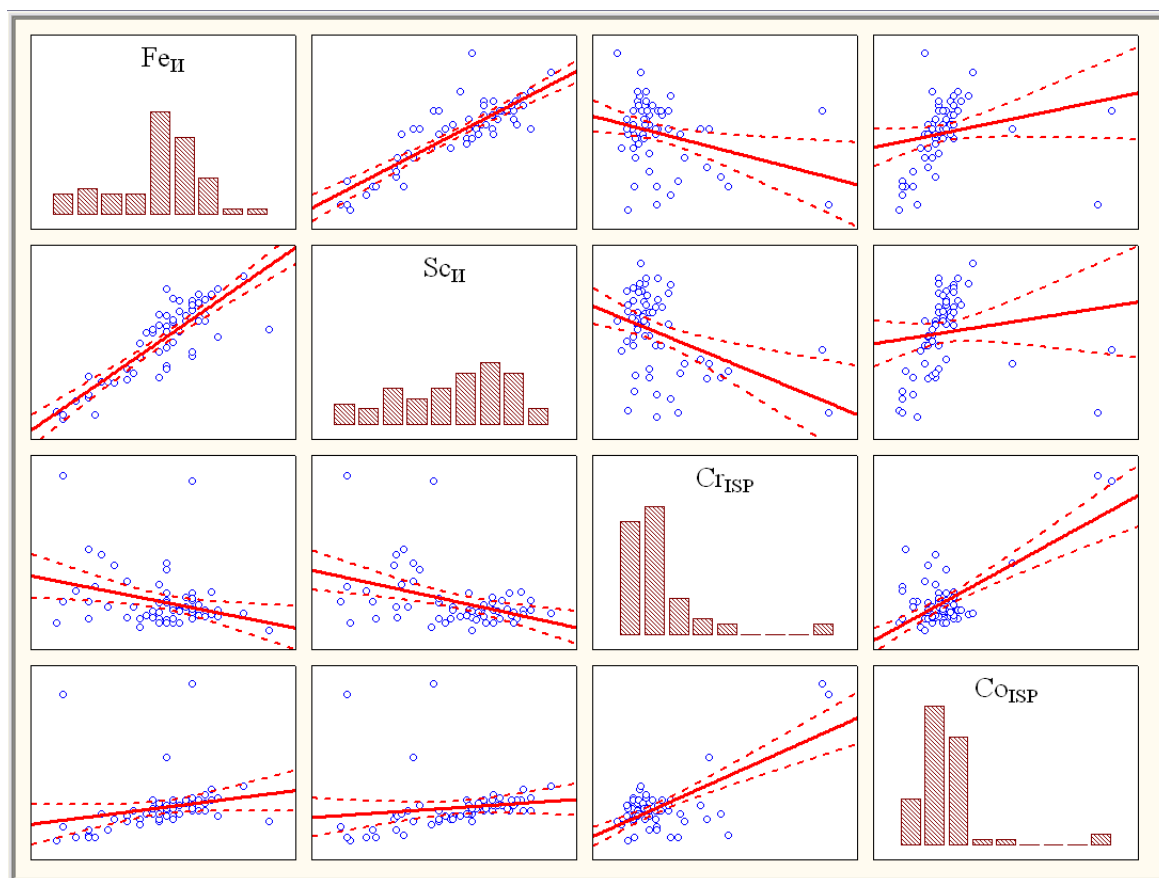

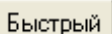
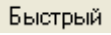





Рис.2.100. Матрица гистограмм и диаграмм рассеяния с прямой линейной регрессии для исследуемых переменных

Подобные графики представляют наглядный и эффективный способ визуального анализа тесноты зависимостей между исследуемыми переменными. На диагонали строятся гистограммы для соответствующих переменных, которые можно использовать для визуальной оценки закона нормального распределения переменной – одного из наиболее важных предположений, выполнение которых обеспечивает получение достоверных и обоснованных результатов.

Для продолжения канонического анализа необходимо в окне

 **Определение модели:** (рис. 2.97) перейти в режим .

Шаг 5. В режиме  окна  **Определение модели** определите микроэлементы ИНАА в первую группу, а микроэлементы ISP - во вторую группу (рис. 2.101), и нажмите  в каждом окне. На экране появится окно  **Результаты канонического анализа:** (рис. 2.102).

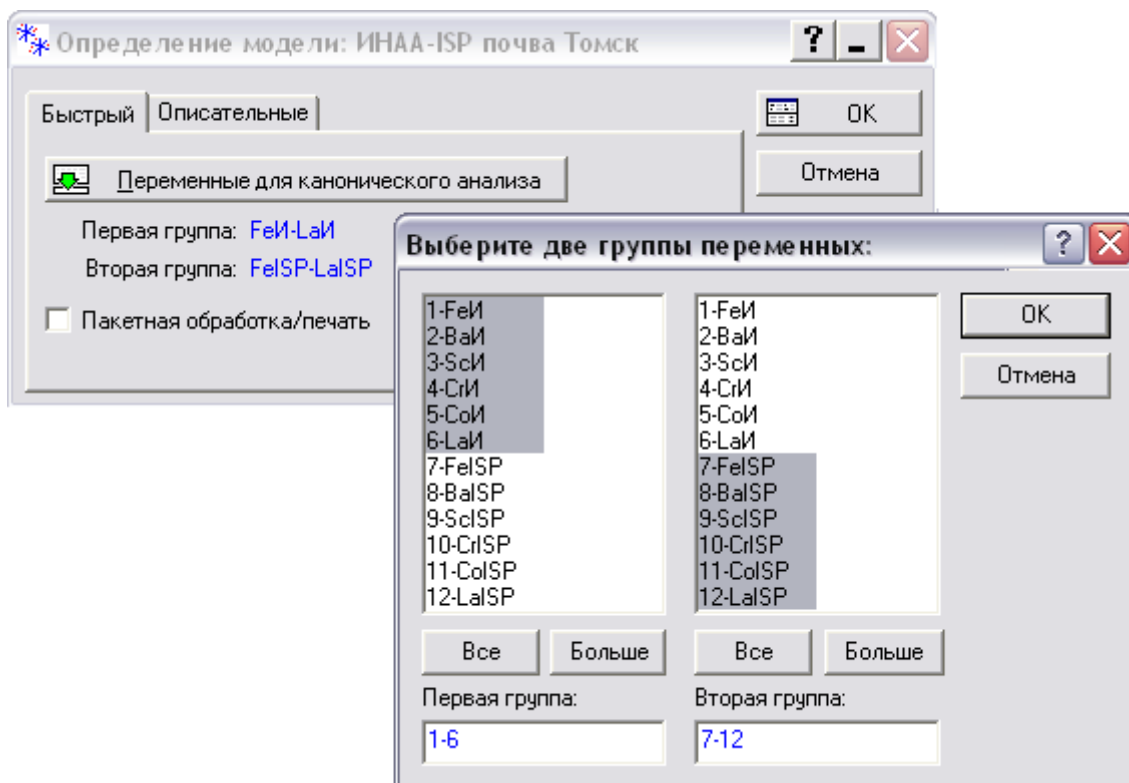


Рис. 2.101. Выбор двух групп переменных для канонического анализа

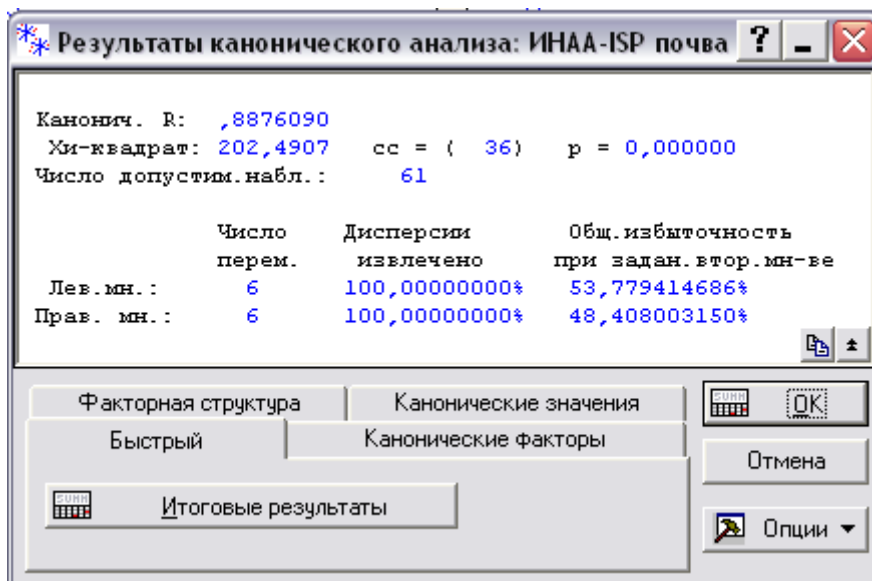


Рис. 2.102. Основные результаты канонического анализа

Окно результатов канонического анализа имеет две части: информационная верхняя часть содержит основную информацию о результатах анализа, а в функциональной нижней части находятся кнопки, позволяющие всесторонне посмотреть результаты анализа.

Согласно рис. 2.102, максимальный канонический коэффициент корреляции $R_1 \approx 0,8876$. Строкой ниже приводится расчетное значение статистики хи – квадрат $\chi^2 \approx 202,49$, которое при степени свободы $ss = 36$ указывает на высоко значимое ($p \approx 0,000000 < 0,0005$) отличие R_1 от нуля. Напомним, что выводимое здесь каноническое значение R относится к первому (наиболее значимому) каноническому корню. Эта величина может быть проинтерпретирована как корреляция между первой парой канонических переменных, то есть между взвешенными суммами соответствующих переменных в первом и втором множестве.

Ниже приведены также строки по группе микроэлементов ИНАА (Лев. мн-во) и группе микроэлементов ISP (Прав. мн-во) для извлеченной дисперсии и общей избыточности, которые являются суммарными характеристиками канонического анализа (п. 1.2.5):

1) *Извлеченная дисперсия.* Коэффициенты канонической корреляции R_k ($k = \overline{1,6}$) соответствуют корреляции между каноническими переменными U_k и V_k , то есть между взвешенными суммами $U_k = \sum_{i=1}^6 a_{ki} X_i$ и

$V_k = \sum_{j=1}^6 b_{kj} Y_j$ по двум множествам переменных X и Y , соответствующи-

ми конкретному каноническому корню λ_k . 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:

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

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

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 6 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 корреляция 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:


$$\frac{1}{6} \sum_{k=1}^6 \sum_{i=1}^6 R_{U_k X_i}^2 R_k^2 \approx 53,78 \% \text{ для Лев. мн-ва и}$$


$$\frac{1}{6} \sum_{k=1}^6 \sum_{j=1}^6 R_{V_k Y_j}^2 R_k^2 \approx 48,41 \% \text{ для Прав. мн-ва.}$$




Полученные значения избыточности можно объяснить следующим образом: пользуясь значениями всех канонических корней и получив значения переменных в правом множестве, можно объяснить, в среднем, 53,78 % дисперсии переменных в левом множестве. Аналогично, можно объяснить 48,41 % изменчивости в правой группе по значениям переменных в левом множестве. Эти результаты говорят о достаточно сильной зависимости между переменными двух множеств.

В окне  Результаты канонического анализа: (рис. 2.102) функциональные клавиши выводят полную информацию с результатами анализа:

 Канонические факторы – итоговые результаты, а также информацию о собственных значениях и их значимости;

 Канонические значения – информацию о канонических весах и значениях канонических переменных для всех наблюдений;






 Факторная структура – информацию о корреляции внутри и между множествами (ассоциациями), о канонических нагрузках и как следствие об извлеченных дисперсиях и избыточности.

Шаг 6. В окне  Результаты канонического анализа: (рис. 2.102) в режиме  Быстрый нажмите  Итоговые результаты для просмотра уже знакомых основных результатов канонического анализа (рис. 2.103).

N=61		Итоги канонич. анализа (ИНАА-ISP почва Томск)	
		Канонич. R: ,88761	
		Chi2 (36)=202,49 p=0,0000	
		Левое Мн-во	Правое Мн-во
Число переменных		6	6
Извлеч. дисперсия		100,000%	100,000%
Общ. избыточность		53,7794%	48,4080%
Переменные :	1	Fe _И	Fe _{ISP}
	2	Ba _И	Ba _{ISP}
	3	Sc _И	Sc _{ISP}
	4	Cr _И	Cr _{ISP}
	5	Co _И	Co _{ISP}
	6	La _И	La _{ISP}

Рис. 2.103. Основные результаты канонического анализа в режиме

 Быстрый окна  Результаты канонического анализа:

Шаг 7. В окне  Результаты канонического анализа: в режиме  Канонические факторы (рис. 2.104) нажмите  Собственные значения для просмотра собственных значений (чисел, корней) λ_k^2 (рис. 2.105) и  График собственных значений для просмотра их графиков, а также нажмите  Chi-квадрат для канонических корней для просмотра результатов оценки значимости коэффициентов канонических корреляций R_k (рис. 2.106).

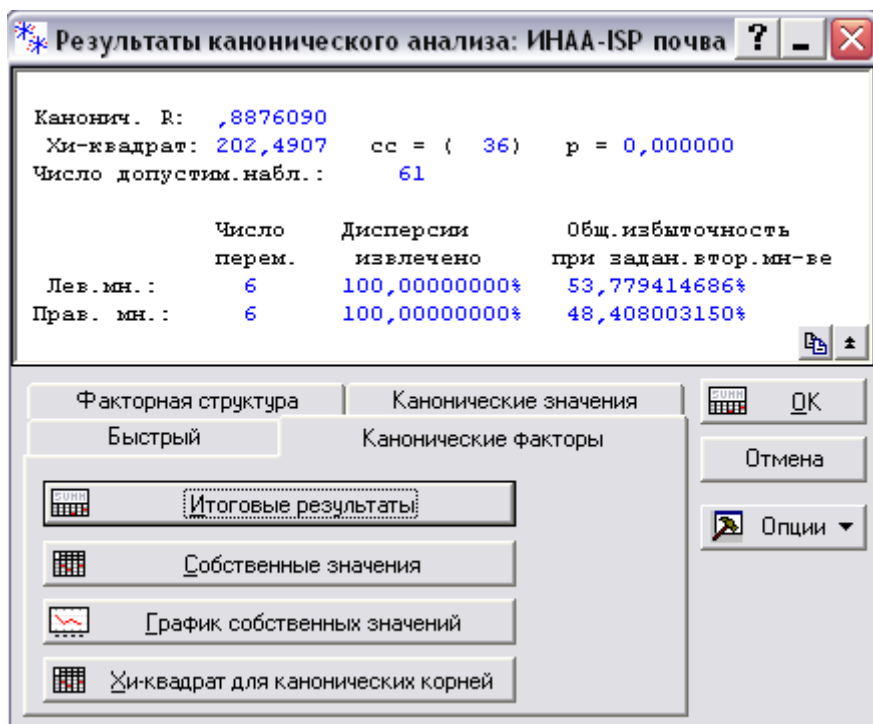


Рис. 2.104. Режим Канонические факторы в окне

Результаты канонического анализа:

Корень	Собств.знач (ИНАА-ISP почва Томск)					
	Кор 1	Кор 2	Кор 3	Кор 4	Кор 5	Кор 6
Знач.	0,7878	0,6713	0,5170	0,2195	0,1052	0,0347

Рис. 2.105. Собственные значения λ_k^2

Корень Удаленный	Значения Хи-квадрат с успешно удален. корнями				
	Канонич. R	Канонич. R-кв.	Хи-кв.	сс	p
0	0,8876	0,7878	202,4907	36	0,000000
1	0,8193	0,6713	119,5411	25	0,000000
2	0,7190	0,5170	60,0238	16	0,000001
3	0,4685	0,2195	21,0926	9	0,012266
4	0,3243	0,1052	7,8325	4	0,097937
5	0,1862	0,0347	1,8878	1	0,169460

Рис. 2.106. Результаты оценки значимости коэффициентов канонических корреляций R_k ($R_k = \lambda_k$)


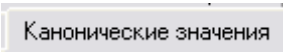
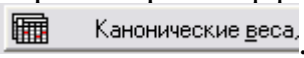
Кроме уже знакомой оценки значимости коэффициента $R_1 \approx 0,8876$ на рис. 2.106 приведена аналогичная оценка значимости коэффициента канонической корреляции $R_2 \approx 0,8193$, для которого значение статистики хи – квадрат $\chi^2 \approx 119,54$ при степени свободы $ss = 25$ указывает высоко значимое ($p \approx 0,000000 < 0,0005$) отличие R_2 от нуля.

Следует заметить, что последовательный критерий значимости работает следующим образом. Сначала рассматриваются все 6 канонические переменные вместе, т.е. без удаления корней. Полученное значение высоко значимо (p -уровень $< 0,00005$). Далее, первый (наиболее значимый) корень "удаляется" и определяется статистическая значимость 5-ти оставшихся корней. Это значение (во второй строке таблицы результатов) также высоко значимо (p -уровень $< 0,00005$). Затем "удаляются" два первых значимых корней и определяется статистическая значимость 4-х оставшихся корней. Это значение (в третьей строке таблицы результатов) также высоко значимо (уровень $p \approx 0,000001 < 0,00005$). При "удалении" трех первых значимых корней определяется значимость 3-х оставшихся корней как статистически значимая ($0,005 < p \approx 0,012 < 0,05$) согласно четвертой строке таблицы результатов (рис. 2.106). На этом можно остановиться, так как последние два корня оцениваются как незначимы ($0,05 < p \approx 0,1$) согласно пятой строке таблицы результатов, и заключить, что только первые четыре корня являются статистически значимыми 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:

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

Канонические веса, вычисленные на основании матрицы парных корреляций, определяются для стандартизированных переменных.

Шаг 8. В окне  "Результаты канонического анализа" в режиме  "Канонические значения" (рис. 2.102) для просмотра коэффициентов a_{ki} (рис. 2.107) и b_{kj} (рис. 2.108) нажмите  "Канонические веса".

Перем	Канон.веса, левое мн-во (ИНАА-ISP почва Томск)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Fe _И	0,931	1,195	1,172	0,731	0,363	-0,244
Ba _И	-0,015	0,253	-0,105	-0,361	-0,292	1,000
Sc _И	0,040	-1,352	-1,089	-0,608	-1,351	-0,192
Cr _И	-0,398	0,156	0,411	-1,635	-0,465	-0,654
Co _И	0,127	0,210	-1,237	1,316	0,573	-0,033
La _И	0,028	0,076	0,034	-0,717	1,200	0,333

Рис. 2.107. Результаты вычисления вектора A_k коэффициентов a_{ki}

(канонических весов) для представления $U_k = \sum_{i=1}^6 a_{ki} X_i$


Перем	Канон.веса, правое мн-во (ИНАА-ISP почва Томск)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Fe _{ISP}	0,960	0,838	1,056	-0,033	0,594	-0,712
Ba _{ISP}	-0,206	0,658	-0,085	0,406	-0,018	1,095
Sc _{ISP}	-0,746	-1,036	-1,011	1,006	0,452	0,831
Cr _{ISP}	-0,527	-0,126	0,334	-1,653	0,898	-0,259
Co _{ISP}	0,312	0,115	-1,135	0,978	-0,652	-0,734
La _{ISP}	0,489	-0,050	0,069	-1,735	-0,126	0,260

Рис. 2.108. Результаты вычисления вектора B_k коэффициентов b_{kj}

(канонических весов) для представления $V_k = \sum_{j=1}^6 b_{kj} Y_j$

В рамках интерпретации канонических корней рассмотрим обычные корреляции R_{UX} и R_{VY} между каноническими переменными и переменными из каждого множества, которые называются каноническими нагрузками факторов и используются при интерпретации канонических переменных. При описании смысла канонической переменной исходят в основном из реального смысла сильно коррелированных переменных, так как считается, что переменные, сильно коррелированные с канонической переменной, имеют с ней много общего.

Шаг 9. В окне  Результаты канонического анализа: в режиме

 Факторная структура

(рис. 2.109) нажмите



Факторная структура и избыточности

— самую нижнюю кнопку для просмотра блока вычислений корреляций

R_{UX} и R_{VY} и компонент извлеченной дисперсии и полной избыточности (рис. 2.110-113).

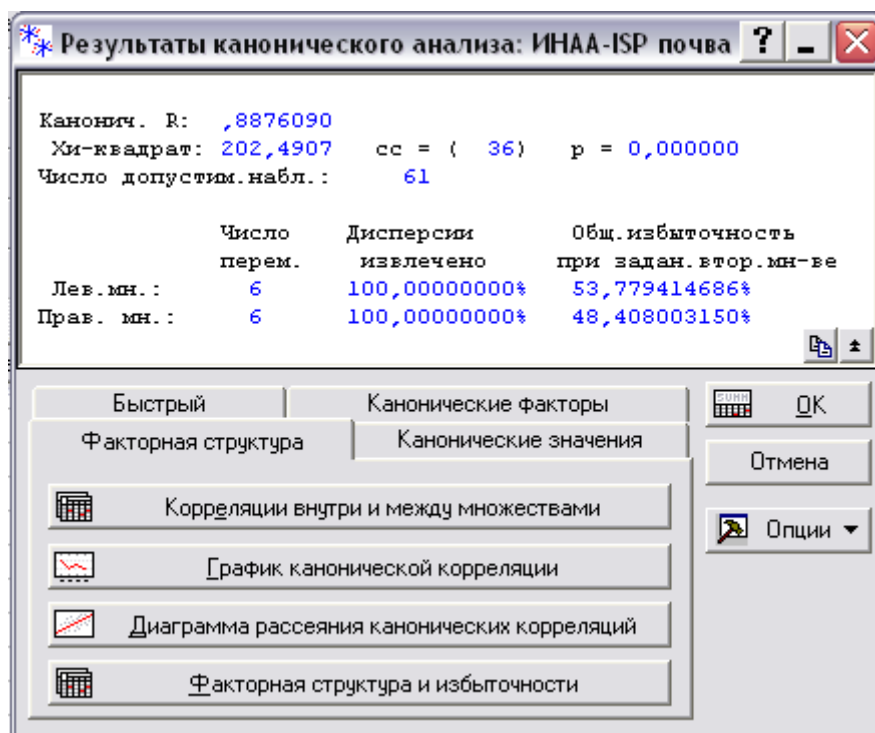


Рис. 2.109. Режим **Факторная структура** в окне

Результаты канонического анализа:

На экране появится одна из четырех таблиц. При этом слева от таблицы отражен столбцом протокол выполненных вычислений (рис. 2.112-113). Достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например, Факт. структ., лев. мн-во – для просмотра таблицы парных корреляций $R_{U_k X_i}$ (рис. 2.110).

Перем	Факт.структ., лев.мн-во (ИНАА-ISP почва Томск)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Fe _и	0,946	0,240	-0,053	-0,145	-0,086	-0,130
Va _и	0,096	0,570	-0,287	-0,208	-0,336	0,654
Sc _и	0,885	-0,257	-0,266	-0,243	-0,122	-0,073
Cr _и	-0,144	0,695	-0,382	-0,399	-0,155	-0,408
Co _и	0,079	0,629	-0,731	-0,049	0,019	-0,248
La _и	0,640	-0,255	-0,153	-0,412	0,572	0,076

Рис. 2.110. Таблица парных корреляций $R_{U_k X_i}$ множества ИНАА

Перем	Факт.структ., прав.мн-во (ИНАА-ISP почва Томск)					
	Корен 1	Корен 2	Корен 3	Корен 4	Корен 5	Корен 6
Fe _{ISP}	0,676	0,087	-0,113	0,136	0,709	0,027
Ba _{ISP}	-0,158	0,814	-0,354	-0,039	0,193	0,384
Sc _{ISP}	0,484	-0,372	-0,340	0,124	0,658	0,253
Cr _{ISP}	-0,506	0,463	-0,387	-0,251	0,475	-0,302
Co _{ISP}	0,109	0,475	-0,792	-0,074	0,204	-0,296
La _{ISP}	0,774	-0,200	-0,356	-0,318	0,075	0,357

Рис. 2.111. Таблица парных корреляций $R_{v_k y_j}$ множества ISP

Фактор	Доли извлеч.дисперсии	
	Извлечен дисперс.	Избыточн
Корен 1	0,354	0,279
Корен 2	0,232	0,156
Корен 3	0,143	0,074
Корен 4	0,076	0,017
Корен 5	0,081	0,009
Корен 6	0,114	0,004

Рис. 2.112. Таблица извлеченной дисперсии и соответствующей избыточности левого множества ИНАА для разных корней

Переменная	Доли извлеч.дисперсии	
	Извлечен дисперс.	Избыточн
Корен 1	0,264	0,208
Корен 2	0,215	0,144
Корен 3	0,193	0,100
Корен 4	0,034	0,008
Корен 5	0,208	0,022
Корен 6	0,086	0,003

Рис. 2.113. Таблица извлеченной дисперсии и соответствующей избыточности правого множества ISP для разных корней

Согласно рис. 2.110, микроэлементы ИНАА Fe_И, Sc_И и La_И имеют значительную нагрузку на первый канонический фактор U₁, а Ba_И, Cr_И и Co_И – на второй канонический фактор U₂. Как видно из рис. 2.112, первый канонический корень извлекает в среднем около 35,4% дисперсии из переменных ИНАА, а второй – 23,2%. Если умножить эти значения на соответствующие доли дисперсии, общие между каноническими переменными в двух множествах (т.е. на R_k^2), то получатся числа во втором столбце таблицы результатов (*избыточность*). Поэтому, задавая значения переменных в правом множестве (факторы ISP), вы можете объяснить соответственно 29,7% и 15,6% дисперсии в переменных ИНАА, исходя из значения соответственно первого и второго канонических корней. Аналогично интерпретируются результаты рис.2.111 и 2.113 для правого множества ISP: Fe_{ISP}, Sc_{ISP} и La_{ISP} имеют значительную нагрузку на первый канонический фактор V₁, а Ba_{ISP}, Cr_{ISP} и Co_{ISP} – на второй канонический фактор V₂.

Заметим, что в режиме Факторная структура (рис. 2.109) можно посмотреть в пакетном режиме аналогично предыдущему случаю блочные корреляционные матрицы исходных переменных (см. рис. 2.99)


Шаг 9'. В окне * Результаты канонического анализа: в режиме Факторная структура (рис. 2.109) нажмите Корреляции внутри и между множествами для построения блочных корреляционных матриц исходных переменных (см. рис. 2.114).





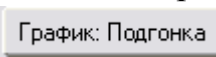
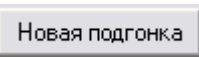
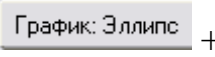
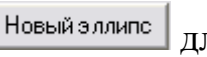

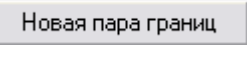
Перем	Корреляции, левое м-во (ИНАА-ISP почва Томск)					
	Fe _И	Ba _И	Sc _И	Cr _И	Co _И	La _И
Fe _И	1,000	0,216	0,845	0,174	0,302	0,552
Ba _И	0,216	1,000	0,058	0,360	0,418	-0,097
Sc _И	0,845	0,058	1,000	-0,059	0,130	0,698
Cr _И				1,000	0,823	-0,167
Co _И				0,823	1,000	0,013
La _И						1,000

Перем	Корреляции, прав. мн-во (ИНАА-ISP почва Томск)					
	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Fe _{ISP}	1,000	0,147	0,824	0,037		
Ba _{ISP}	0,147	1,000	-0,039	0,579		
Sc _{ISP}	0,824	-0,039	1,000	-0,080		
Cr _{ISP}	0,037	0,579	-0,080	1,000		
Co _{ISP}	0,331	0,579	0,195	0,676		
La _{ISP}	0,566	0,005	0,670	-0,338		

Перем	Корреляции между множествами (ИНАА-ISP п					
	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Fe _И	0,559	0,029	0,313	-0,308	0,221	0,635
Ba _И	0,034	0,469	-0,115	0,189	0,344	0,112
Sc _И	0,491	-0,236	0,480	-0,407	0,141	0,747
Cr _И	-0,069	0,550	-0,256	0,481	0,500	-0,087
Co _И	0,152	0,579	0,010	0,429	0,685	0,130
La _И	0,484	-0,172	0,492	-0,209	0,097	0,601

Рис. 2.114. Блочные корреляционные матрицы исходных переменных

Для визуализации результатов канонического анализа можно построить кусочно-линейный график канонических корреляций с помощью кнопки  График канонической корреляции и диаграмму рассеяния пары полученных канонических переменных.

Шаг 9''. В окне  Результаты канонического анализа: (рис. 2.109) в режиме  Факторная структура нажмите  Диаграмма рассеяния канонических корреляций, выберите пару канонических переменных из разных множеств для построения диаграммы рассеяния (см. рис. 2.115). В данном случае добавлены прямая регрессии, 95% эллипс рассеяния и границы 95% доверительного интервала. Для этого аналогично построению рис.2.85-2.86 щелкнуть дважды мышкой в поле диаграммы, в открывшемся окне  Все параметры графика нажать поочередно  График: Подгонка +  Новая подгонка для добавления прямой регрессии,  График: Эллипс +  Новый эллипс для добавления 95% эллипс рассеяния и  График: Регрес. границы +  Новая пара границ для добавления границ 95% доверительного интервала.

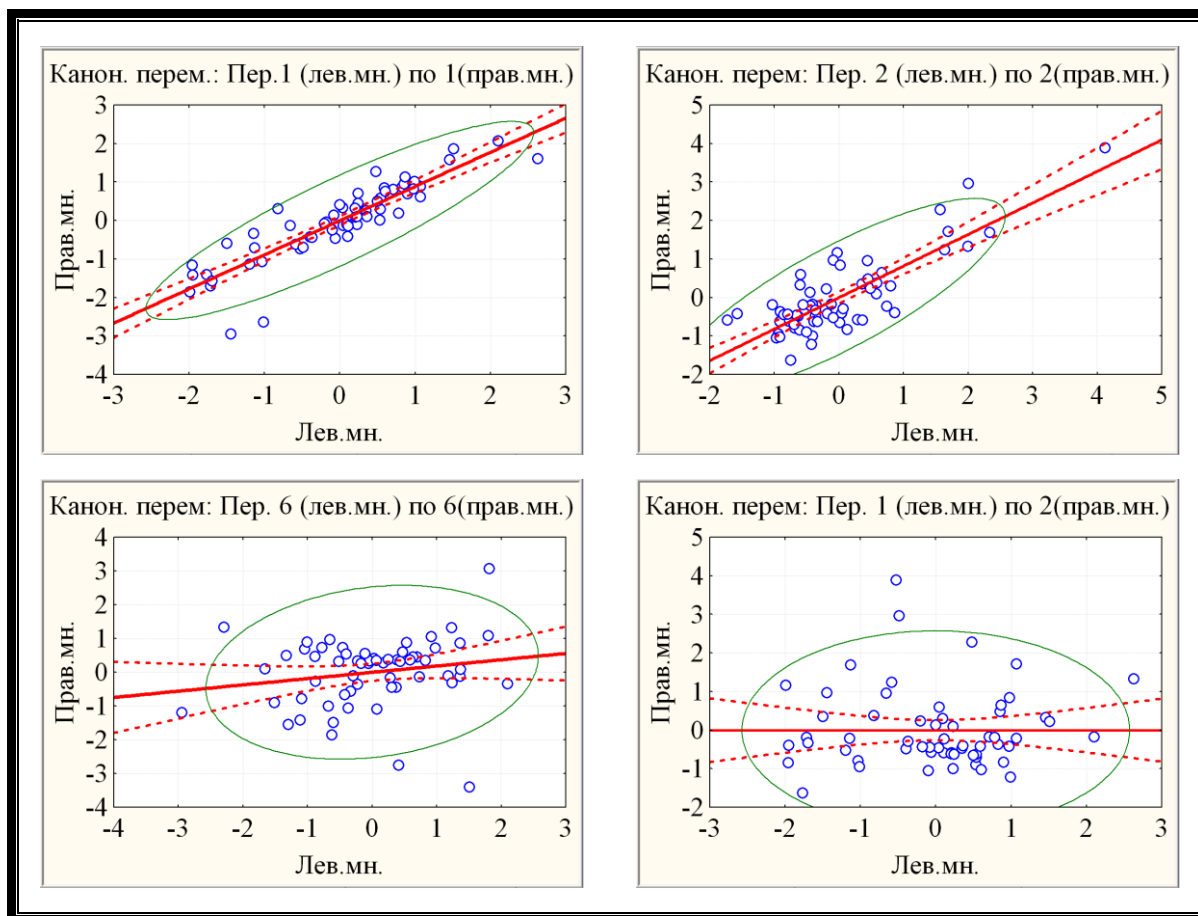


Рис. 2.115. Диаграммы рассеяния канонических переменных.

Построенные диаграммы рассеяния (рис. 2.115) наглядно иллюстрируют основные свойства канонических корреляций: последовательное убывание корреляционной связи между парами канонических переменных, соответствующих одному собственному значению, и не коррелированность между каноническими переменными, соответствующими разным собственным значениям.

Таким образом, можно заключить, что значительная корреляция между микроэлементами в двух ассоциациях (на основании сведений о первых двух корнях), вероятно, является следствием корреляционной зависимости между микро ассоциациями $Fe_{II} - Sc_{II} - La_{II}$ и $Fe_{ISP} - Sc_{ISP} - La_{ISP}$ (первый корень), а также $Ba_{II} - Cr_{II} - Co_{II}$ и $Ba_{ISP} - Cr_{ISP} - Co_{ISP}$ (второй корень).

Заметим, что проведение канонического анализа в STATISTICA возможно в усеченном компактном режиме «Пакетная обработка», который можно заказать на стадии выбора переменных для канонического анализа (см. рис. 2.101).

Шаг 10. Выделите галочкой Пакетная обработка/печать в режиме Быстрый окна **Определение модели:** (2.101') и нажмите .

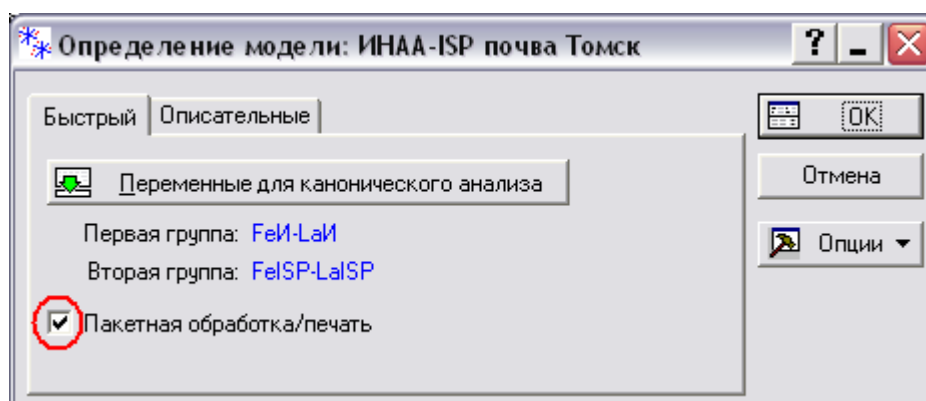


Рис. 2.101'. Выбор двух групп переменных для канонического анализа и пакетного режима обработки

На экране появится таблица собственных чисел (рис.2.105). При этом слева от таблицы отражен столбцом протокол пакетного режима выполненных вычислений (рис. 2.116). Достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например, ... – для просмотра оценок значимости канонических корреляций R_k (рис. 2.106 и 2.116).

... – рис. 2.107, ... – рис. 2.108,
 ... – рис. 2.110, ... – рис. 2.111,
 ... – рис. 2.112, ... – рис. 2.113.

Корень Удален	Значения Хи-квадрат с успешно удален. к	Канонич. R	Канонич. R-кв.	Хи-кв.	сс	р
0		0,8876	0,7878	202,49	36	0,00000
1		0,8193	0,6713	119,54	25	0,00000
2		0,7190	0,5170	60,02	16	0,00000
3		0,4685	0,2195	21,09	9	0,01227
4		0,3243	0,1052	7,83	4	0,09794
5		0,1862	0,0347	1,89	1	0,16946

Рис. 2.116. Результаты оценки значимости коэффициентов канонических корреляций R_k (рис.2.106)



2.6. Кластерный анализ ассоциации микроэлементов

Рассмотрим применения кластерного анализа для сопоставления результатов исследований по методам. Задача кластерного анализа сводится к разбиению множества химических элементов на группы, в которые объединяются элементы с наивысшими значениями меры сходства (парных коэффициентов корреляции Пирсона r). Сначала выявляют наивысший коэффициент корреляции между отдельными парами, затем соответствующую пару объединяют в группу с помощью различных методов осреднения. Такой метод парного объединения применяется до тех пор, пока значения групповых коэффициентов корреляции не достигнут критического порогового значения r_{kp} , определяемого с помощью системы STATISTICA по заданному уровню значимости и объему выборки (см. рис.2.71-72).

Графические результаты кластерного 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.

На **этапе V** для оценки сходства результатов, полученных разными методами, на основе кластерного анализа (метода иерархической группировки) необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в головном меню модуль .

Шаг 2. Выберите  Многомерный разведочный анализ в стартовой панели модуля и далее щелкните мышкой на разновидности  Кластерный анализ (рис.2.117).

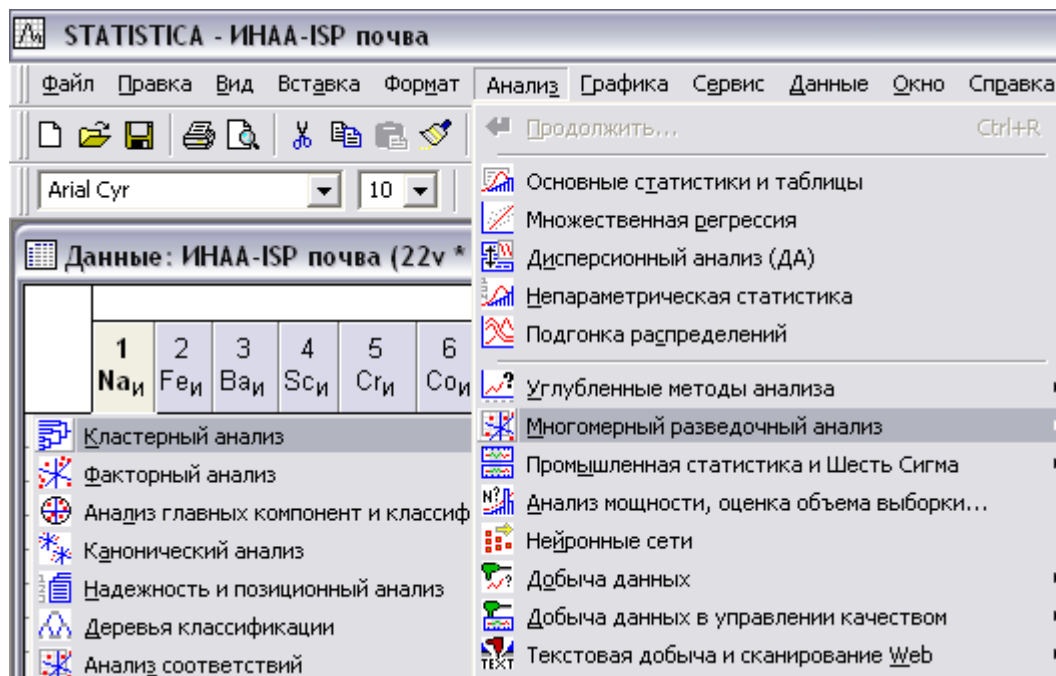






Рис. 2.117. Выбор  Кластерный анализ в модуле 

Шаг 3. В появившемся окне выделите разновидность  Иерархическая классификация и щелкните  (рис.2. 118).

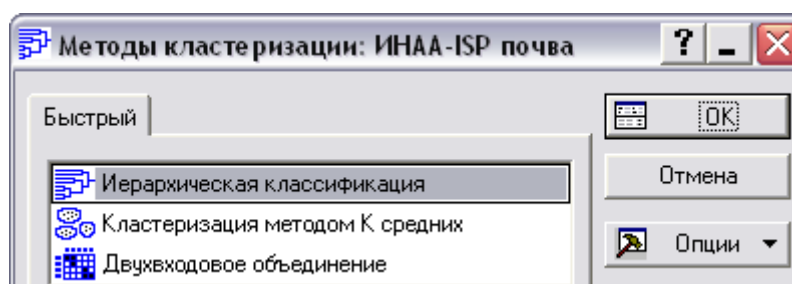


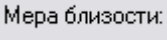

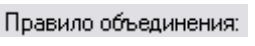




Рис. 2.118. Окно  Методы кластеризации:

Шаг 4. В открывшемся окне нажмите кнопку  Переменные: и выберите все переменные базы данных метода ИНАА и нажмите .

Шаг 5. Далее в окне  Мера близости: выделите  1-г Пирсона, а в окне  Правило объединения: —  Метод полной связи и нажмите  (рис.2.119).

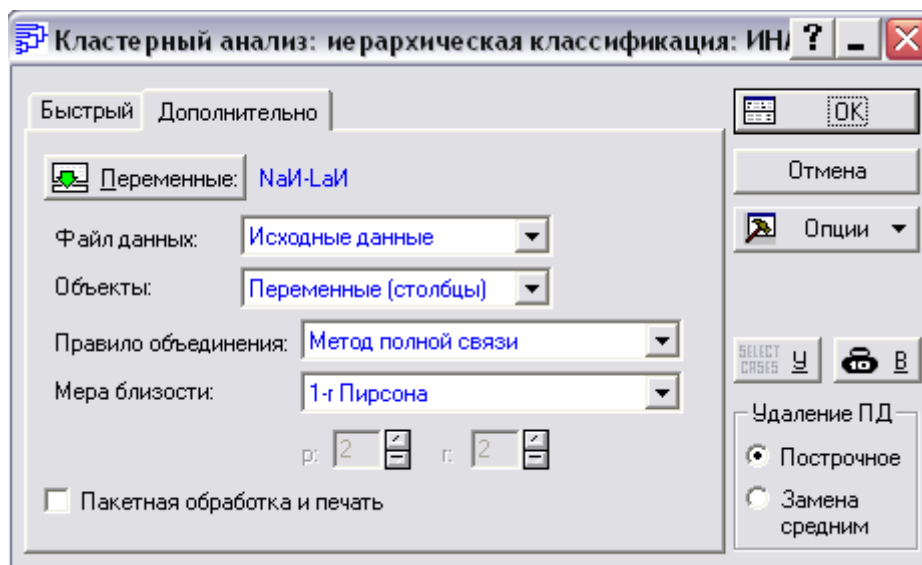


Рис. 2.119. Окно **Кластерный анализ** для переменных

Шаг 6. После этого щелкните окно **Вертикальная дендрограмма** (рис.2.120).

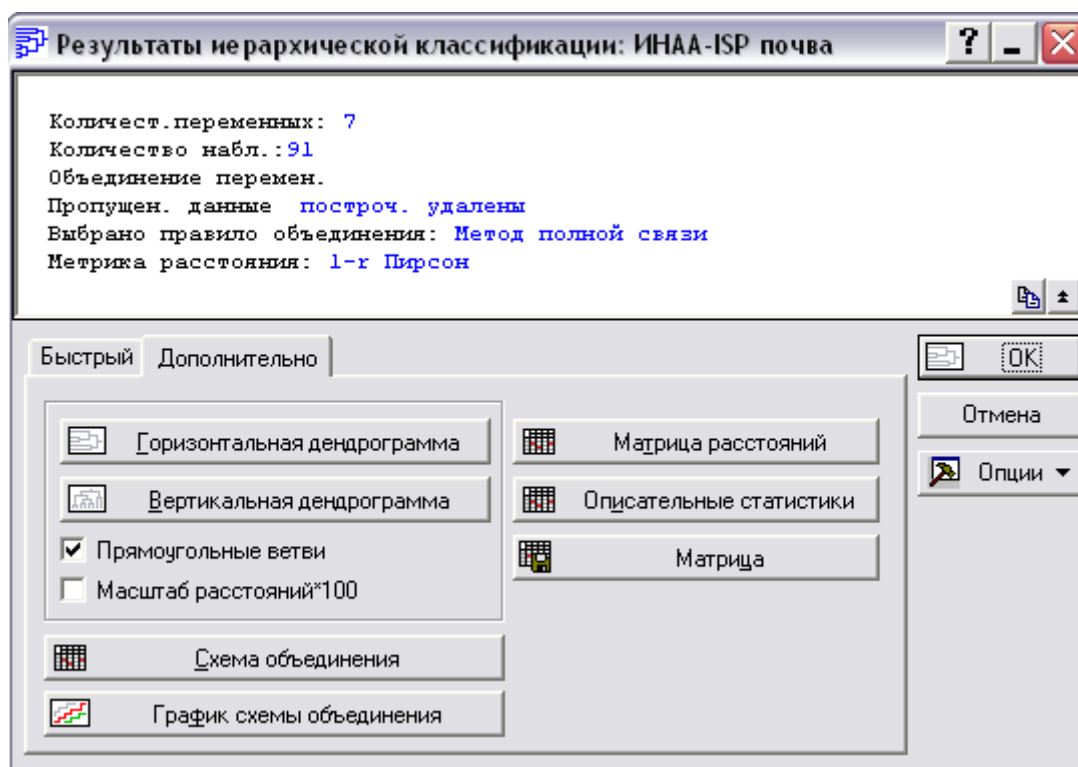


Рис. 2.120. Окно **Результаты нерархической классификации**.
Следующая дендрограмма по методу ИНАА на экране (рис.2.121).

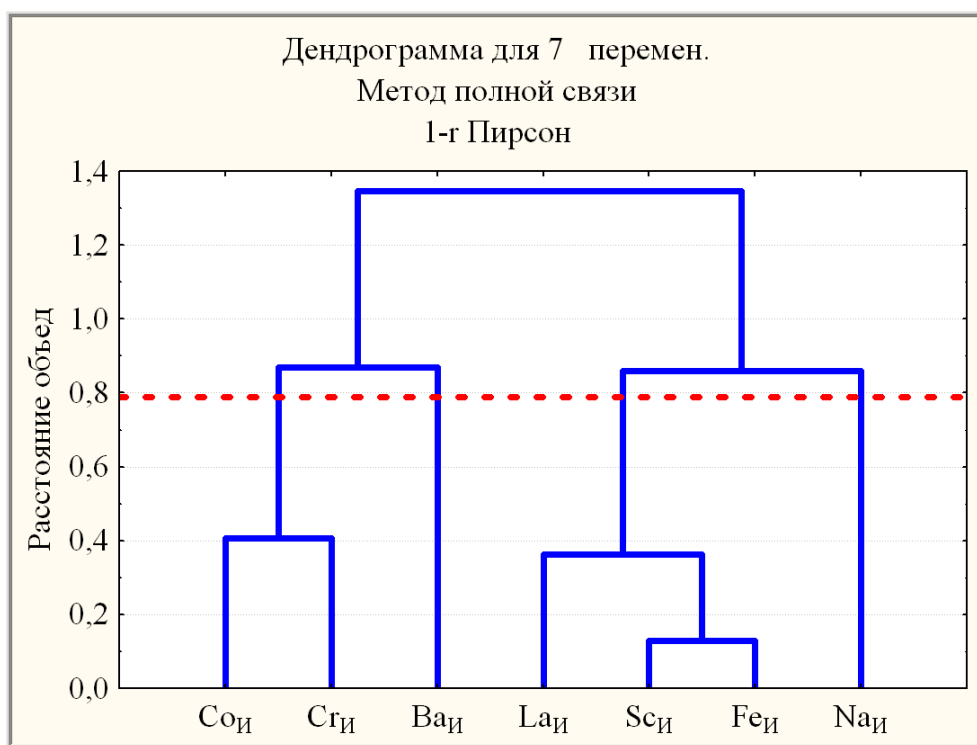


Рис. 2.121. Дендрограмма корреляционной матрицы геохимического спектра микроэлементов по методу ИНАА ($N = 91$; $1 - r_{0,05} = 0,79$)

Аналогично – по методу ISP (рис.2.122)

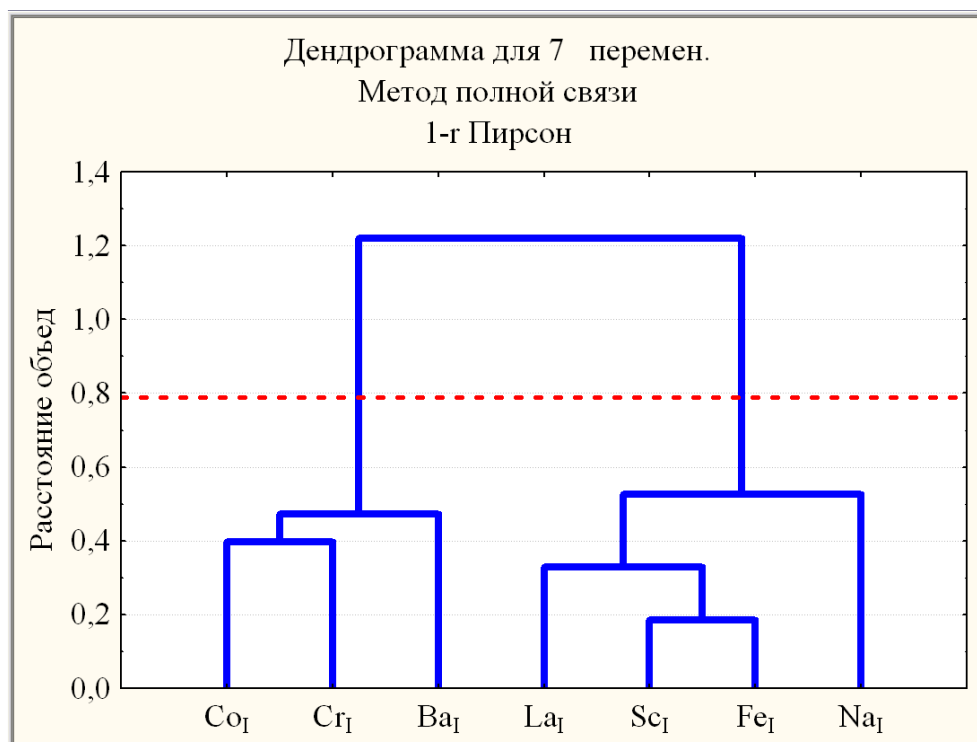


Рис. 2.122. Дендрограмма корреляционной матрицы геохимического спектра микроэлементов по методу ISP ($N = 91$; $1 - r_{0,05} = 0,79$).

Как следует из рис.2.121 - 122 геохимические спектры микроэлементов в почве образуют разные по методам значимые микроассоциации химических элементов: {Fe – Sc – La; Co – Cr} по методу ИНАА и {Fe – Sc – La – Na; Co – Cr – Ba} по методу ISP.

Для изображения критического уровня ($1 - r_{0,05} = 0,79$ на рис.2. 121 - 122) необходимо выполнить последовательность следующих действий:

Шаг 1. Дважды щелкнув мышкой по дендрограмме, войти в модуль **Все параметры графика**.

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

Шаг 3. Заказать ряд параметров в окнах справа: в окне **Тип:** - тип параметрической кривой, в окне **Браузер функций** - задание параметрической функции $\{Y(t) = t, X(t) = 0.79\}$, в окне **Линия** с помощью кнопки **Шаблон...** - выбрать Ширину, Шаблон (пунктирный) и Цвет (красный) линии, в окне **Диапазон изменения параметра t** - значения 0 – 8, где число 8 - число химических элементов плюс 1 (рис.2.123)

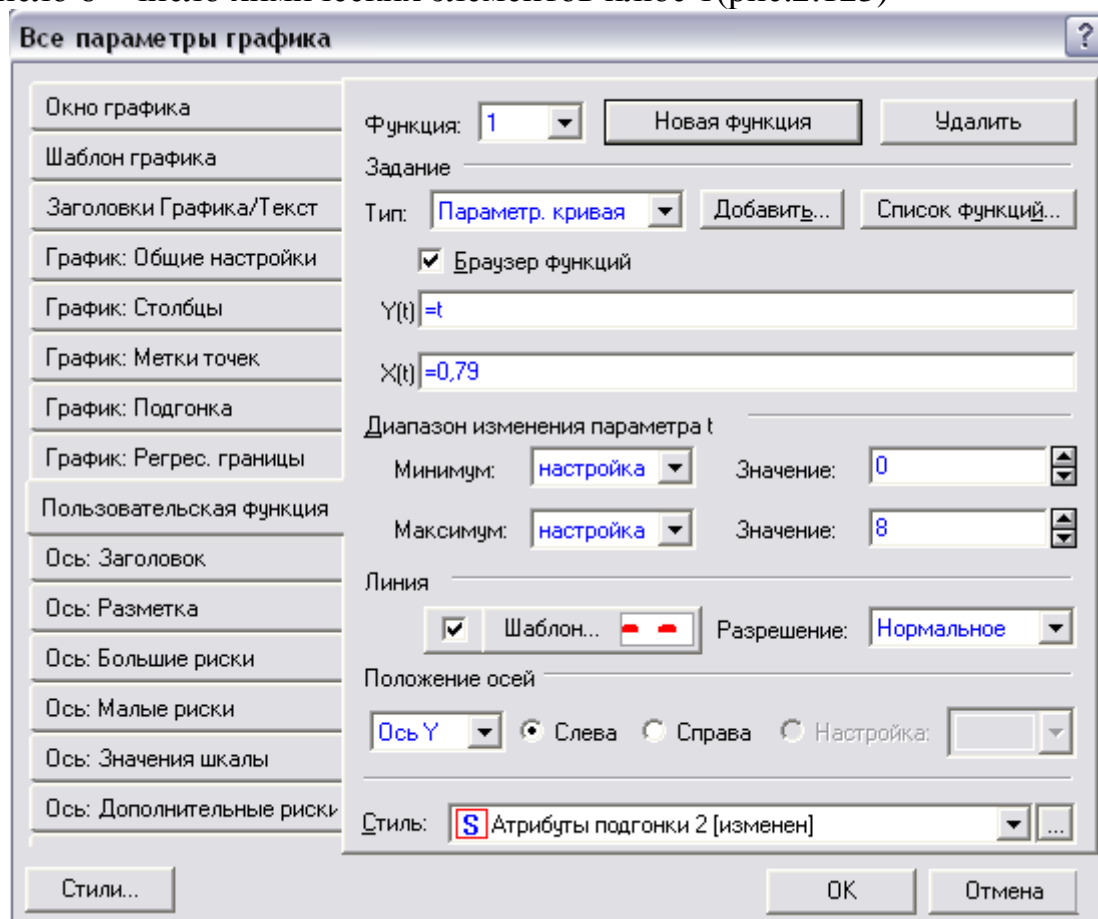


Рис. 2.123. Режим **Пользовательская функция** в модуле **Все параметры графика**

На этапе VI проведение кластеризации наблюдений содержаний химических элементов предполагает:

1. проведение факторного анализа содержаний химических элементов и построение оптимальной факторной модели содержаний,
2. проведение кластерного анализа наблюдений в рамках построенной факторной модели содержаний.

Рассмотрение проведем на примере содержаний химических элементов, полученных методом ISP, в почве территории Томского региона юго-западного (ЮЗ) сектора влияния Сибирского химического комбината (измерения на территории с. Зоркальцево №1–19 и с. Поросино №81–83 в приложении П₁). Соответствующий фрагмент выделен из П₁ в файл с названием ISP_ЮЗ_почва.sta (рис.2.124).

	1	2	3	4	5	6	7
	Na _{ISP}	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
1	2,16	3,16	370	10,2	144,5	21,1	28
2	2,27	3,22	350	9,1	161,2	21,6	26,2
3	2,32	3,61	330	10	127,7	23	24,9
4	2,54	3,82	380	13,4	136,7	22,6	33,1
5	2,46	3,26	320	9,7	136,4	22,1	27,5
6	1,99	2,78	420	10,2	213,5	16,8	26,2
7	2,47	2,99	380	11,2	73,5	20,9	27
8	2,58	3,28	370	11,3	76,7	24	27,8
9	2,62	3,62	350	11	173	22,8	26,4
10	2,23	2,36	430	9,8	103,5	17,2	27,6
11	2,8	3,03	430	11,8	62,9	24,8	28,1
12	2,76	4	400	13,5	125	23,1	27,3
13	2,79	3,88	390	13,4	79,2	22,8	26,6
14	2,92	4,18	400	14,3	148,4	24	31,1
15	2,21	2,54	410	10	64,9	19,9	30,8
16	2,45	2,58	390	10,9	64,8	19,8	29,7
17	2,96	3,72	390	13,6	128	23	28,9
18	2,6	3,85	380	14	115,4	23,3	31,4
19	2,78	3,97	380	14	116,3	23,1	30,2
20	1,52	2,97	380	10	100	11	18
21	1,83	3,07	350	11	90	13	20
22	1,99	3,79	340	14	100	14	19

Рис. 2.124. Содержания химических элементов ISP_ЮЗ_почва

Построение корреляционной матрицы данной ассоциации микро-элементов можно выполнить аналогично рис. 2.69-2.70 и получить табл. 2.9, в которой значимые, то есть отличные от нуля на уровне значимости 0,05, коэффициенты парной корреляции Пирсона выделены

жирным шрифтом. Значение критического коэффициента корреляции Пирсона по заданному уровню значимости 0,05 и объему выборки (22 наблюдения) вычисляется аналогично рис. 2.71-2.72 и равно $r_{кр} \approx 0,423$.

Таблица 2.9

Корреляционная матрица ассоциации микроэлементов в почво-грунтах ЮЗ по методу ISP.

	Na _{ISP}	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Na _{ISP}	1,00	0,55	0,21	0,58	-0,04	0,89	0,71
Fe _{ISP}	0,55	1,00	-0,30	0,79	0,26	0,45	0,15
Ba _{ISP}	0,21	-0,30	1,00	0,15	-0,17	0,07	0,38
Sc _{ISP}	0,58	0,79	0,15	1,00	-0,08	0,31	0,27
Cr _{ISP}	-0,04	0,26	-0,17	-0,08	1,00	0,06	0,06
Co _{ISP}	0,89	0,45	0,07	0,31	0,06	1,00	0,77
La _{ISP}	0,71	0,15	0,38	0,27	0,06	0,77	1,00

Там же можно построить матричную диаграмму рассеяния с гистограммами выбранных переменных, например, наиболее коррелирующих микроэлементов Na_{ISP}, Co_{ISP} и La_{ISP} (рис.2.125).

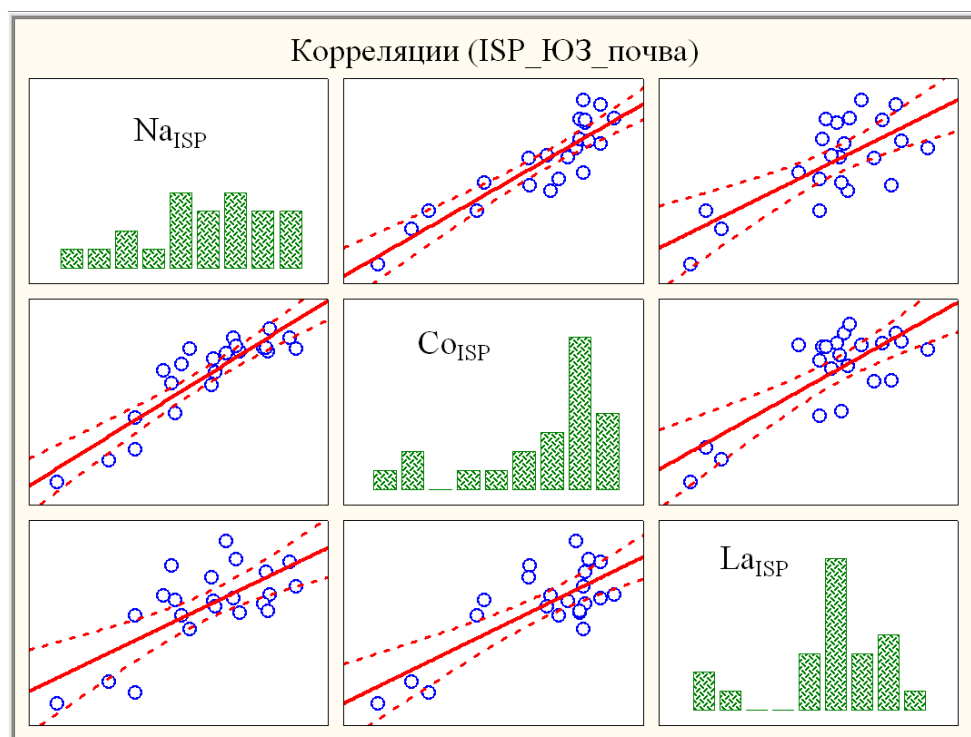


Рис. 2.125. Матричная диаграмма рассеяния с гистограммами микроэлементов Na_{ISP}, Co_{ISP} и La_{ISP}

Можно заметить, что результаты корреляционного анализа ISP_ЮЗ по 22 наблюдениям (табл. 2.9) значительно отличаются от соответствующих результатов (табл. 2.5) по 91 наблюдению.

Поскольку корреляционный анализ данной ассоциации микроэлементов показал значимые корреляционные связи, то, при проведении кластеризации микроэлементов, целесообразно использовать корреляционное расстояние в качестве меры близости микроэлементов. Графические результаты кластерного анализа микроэлементов ISP_ЮЗ представлены на дендрограммах (рис.2.126), построенных разными методами объединения кластеров аналогично рис. 2.117 - 2.120.

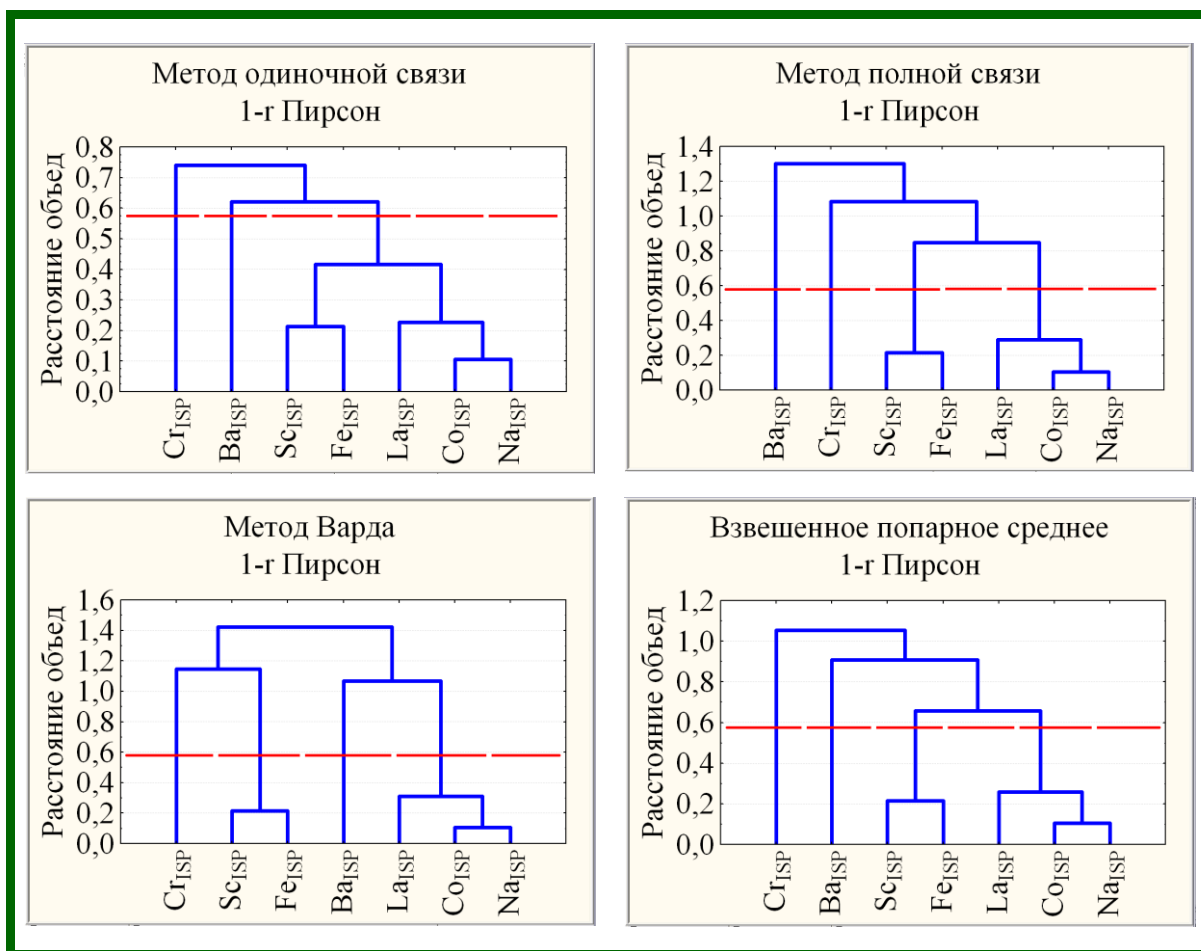


Рис. 2.126. Дендрограммы корреляционной матрицы геохимического спектра микроэлементов по методу ISP ($N = 22$; $1 - r_{0,05} = 0,577$).

Преимуществом кластерного анализа является то, что он позволяет выявлять не только высоко значимые парные корреляции микроэлементов, но и значимые ($1 - r \leq 1 - r_{кр} = 1 - 0,423 = 0,577$ - критическое значение для группы из 22 наблюдений при уровне значимости 0,05 выделено на рис.2.126 горизонтальной пунктирной прямой, по-

строенной аналогично рис.2.123) групповые кластеры микроэлементов $\{Na_{ISP} - Co_{ISP} - La_{ISP}, Fe_{ISP} - Sc_{ISP}, Ba_{ISP}, Cr_{ISP}\}$, т.е. факторы для построения факторной модели ассоциации микроэлементов. Корреляционная зависимость между микроэлементами внутри групповых кластеров (табл. 2.9) оценивается аналогично рис. 2.71-2.72 и является высоко значимой (например, при $N = 22$ по $r = 0,71$ между Na_{ISP} и La_{ISP} вычисления дают уровень значимости $p \approx 0,0002 < 0,0005$).

Заметим, что дендрограммы корреляционной матрицы ISP_ЮЗ по 22 наблюдениям (рис. 2.126) существенно отличаются от соответствующей дендрограммы (рис. 2.122) по 91 наблюдению.

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.

2.6. Факторный анализ ассоциации микроэлементов

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. Откройте файл ISP_ЮЗ_почва.sta (рис.2.124).

Шаг 2. Запустите **Анализ**, нажмите **Многомерный разведочный анализ**, где далее выберите **Факторный анализ** (рис. 2.127).

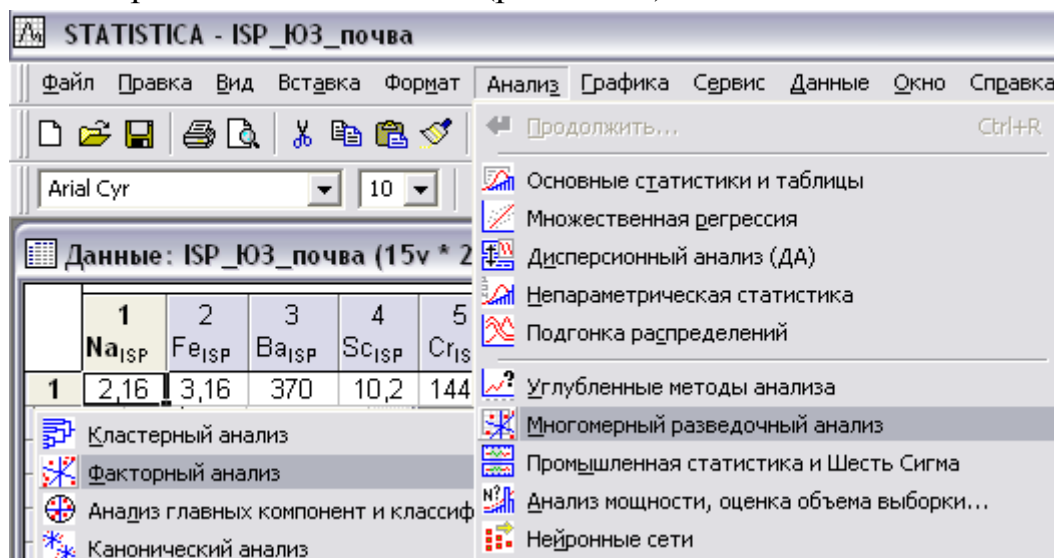




Рис. 2.127. Модуль **Факторный анализ** в пакете STATISTICA

Шаг 3. Выберите  Переменные:  в каждом окне. В результате появится рис. 2.129.

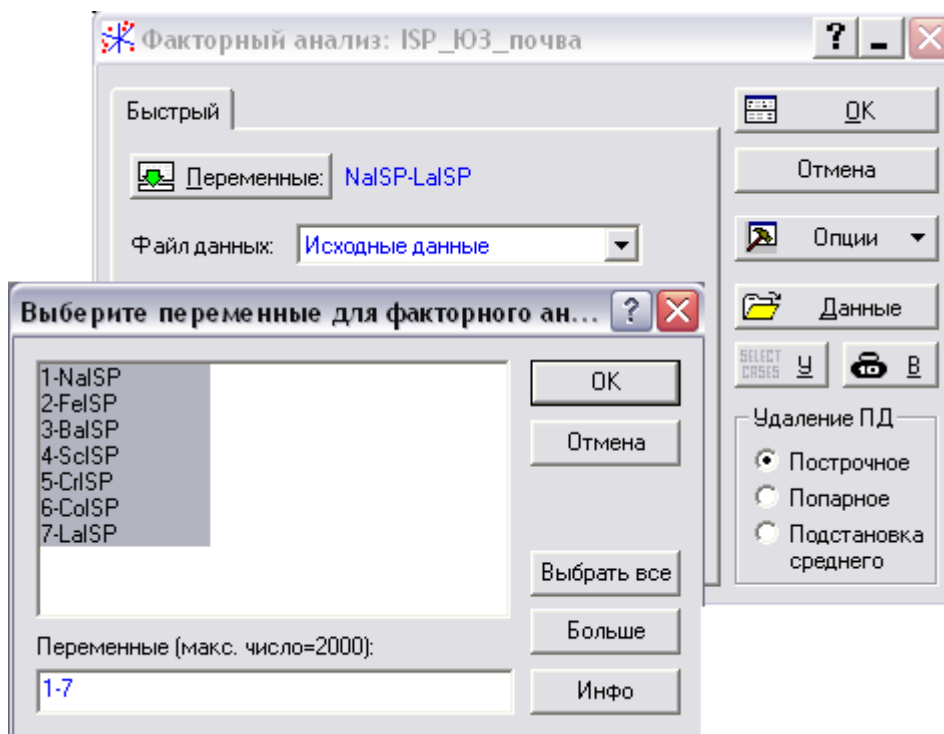


Рис. 2.128. Выбор переменных для факторного анализа

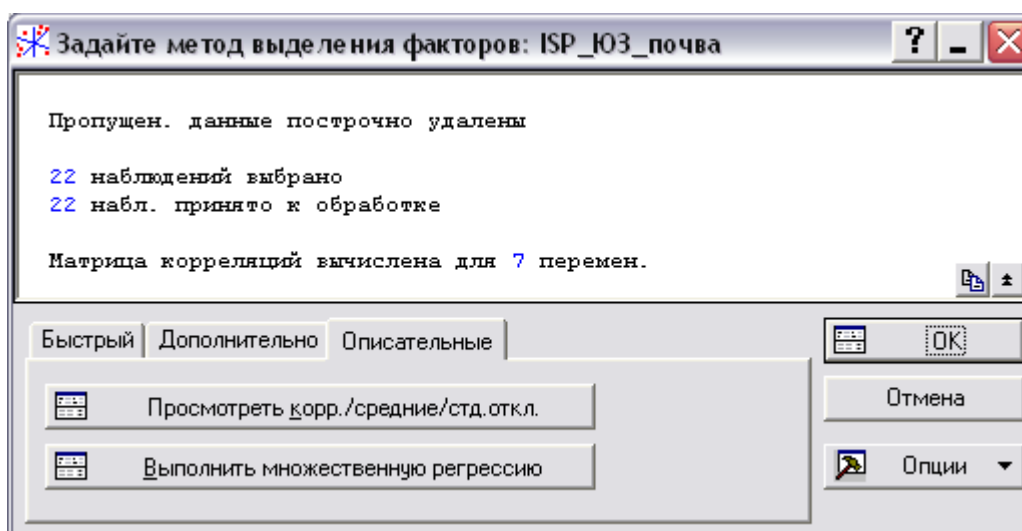
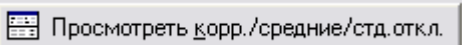
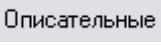


Рис. 2.129. Выбор режима  для просмотра описательной статистики

Шаг 4. Нажмите кнопку  в режиме . В результате появится рис. 2.130.

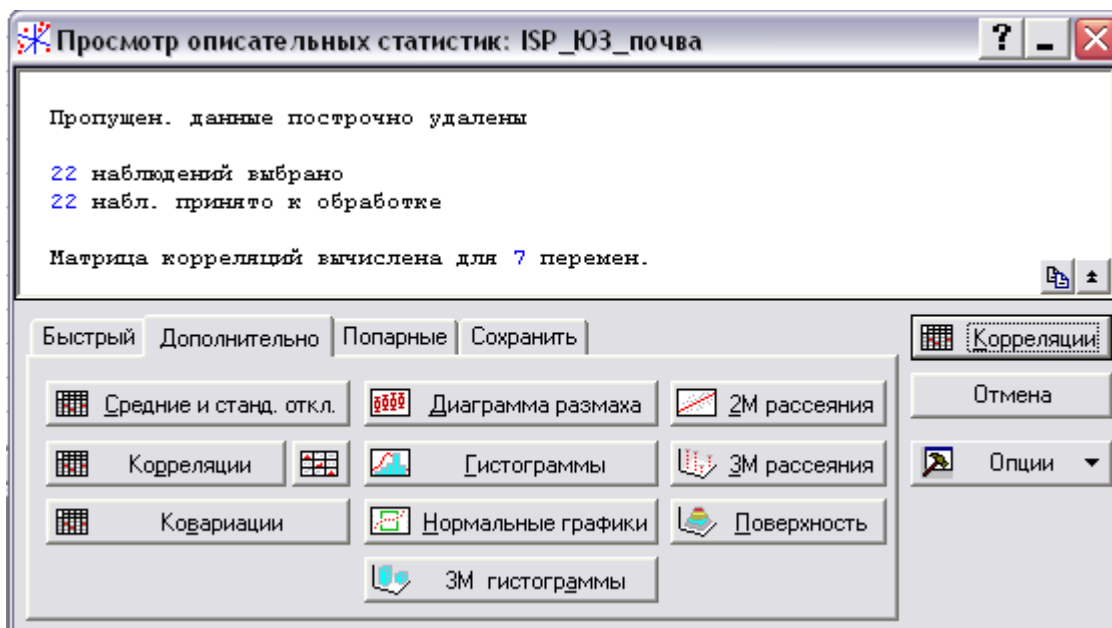


Рис. 2.130. Окно для просмотра описательной статистики в режиме

Дополнительно

Можно посмотреть (рис. 2.130) в режиме **Дополнительно** уже знакомую корреляционную матрицу **Корреляции** (см. табл.2.9) и матричную диаграмму рассеяния с гистограммами **Корреляции** (рис. 2.125), а также прочую описательную статистику выборки (**Средние и станд. откл.** и **Ковариации**), построить различные графики, например, **2М рассеяния** Fe – Sc (рис. 2.131) или **3М рассеяния** Na – Co – La (рис. 2.132).

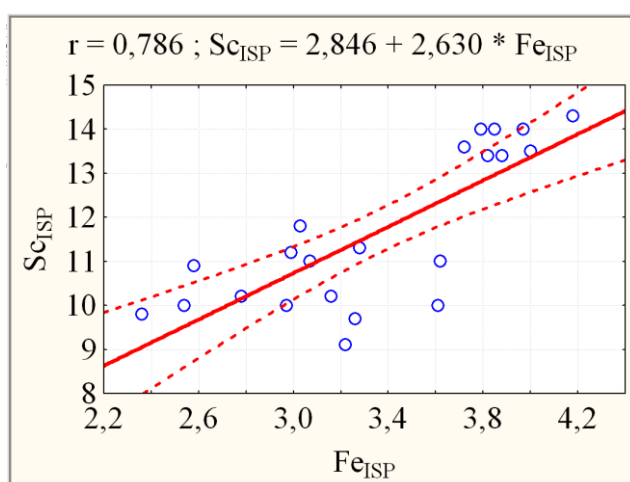


Рис. 2.131. Диаграмма рассеяния Fe_{ISP} – Sc_{ISP}

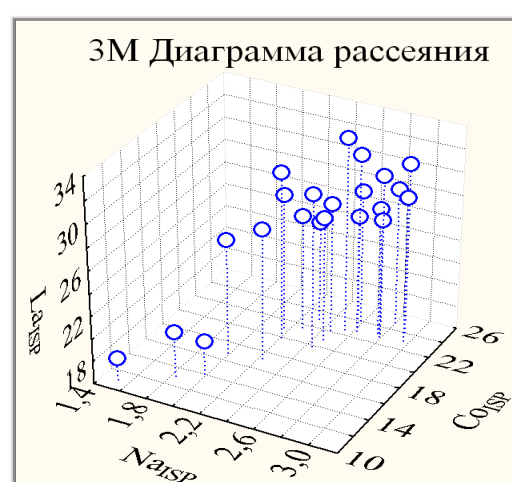


Рис. 2.132. Диаграмма рассеяния Na_{ISP} – Co_{ISP} – La_{ISP}

Описательную статистику можно посмотреть также в режиме

Сохранить (рис. 2.133)

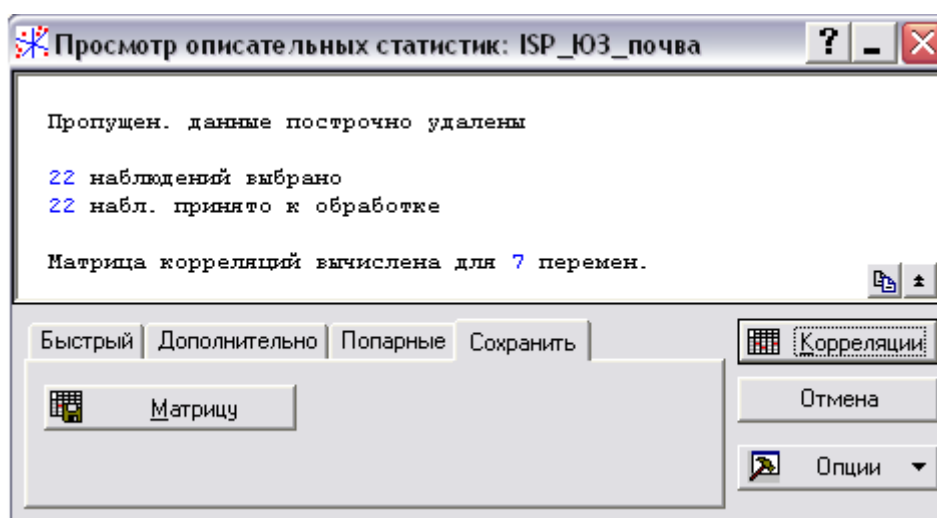


Рис. 2.133. Окно для просмотра описательной статистики в режиме

Шаг 5. В режиме Сохранить (рис. 2.133) нажмите кнопку

Матрицу для вывода на экран корреляционной матрицы, расширенной строками средних и стандартных отклонений (рис. 2.134).

	ISP_Ю3_почва						
	1	2	3	4	5	6	7
	Na _{ISP}	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Na _{ISP}	1,000	0,548	0,209	0,583	-0,036	0,894	0,712
Fe _{ISP}	0,548	1,000	-0,300	0,786	0,259	0,451	0,152
Ba _{ISP}	0,209	-0,300	1,000	0,149	-0,170	0,070	0,379
Sc _{ISP}	0,583	0,786	0,149	1,000	-0,084	0,313	0,272
Cr _{ISP}	-0,036	0,259	-0,170	-0,084	1,000	0,064	0,060
Co _{ISP}	0,894	0,451	0,070	0,313	0,064	1,000	0,772
La _{ISP}	0,712	0,152	0,379	0,272	0,060	0,772	1,000
Средние	2,420	3,349	379,091	11,655	115,527	20,632	27,082
Стд. откл.	0,371	0,525	30,223	1,757	39,014	3,842	3,858
Кол-во N	22						

Рис. 2.134. Наблюдаемая корреляционная матрица, расширенная строками средних и стандартных отклонений выборки ISP_Ю3_почва

Шаг 6. Для проведения факторного анализа выборки вернитесь в окно **Задайте метод выделения факторов** (рис.2.129) с помощью кнопки **Отмена** и выберите режим **Дополнительно** (рис. 2.135), в котором в поле справа можно указать Максимальное число факторов, которые будут выделены системой, например, в начале исследования 7, и Минимальное собственное значение, например 0,00 (рис. 2.135).

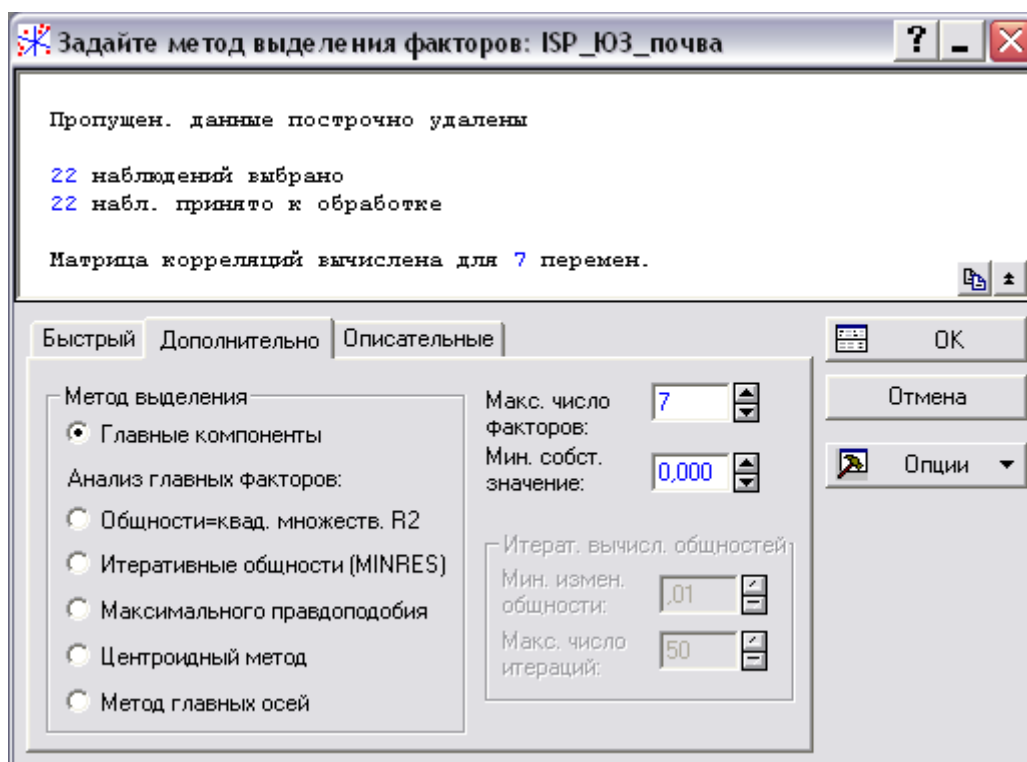


Рис. 2.135. Режим **Дополнительно** для выбора метода анализа

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. 2.136).

В верхней части окна **Результаты факторного анализа** дается информационное сообщение, в частности, приводятся значения собственных значений: 3,26346 1,56983 1,08907 0,775120 0,213418....

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

Шаг 7. Например, для просмотра расширенной информации о собственных значениях в режиме **Объясненная дисперсия** (рис. 2.136) щелкните по кнопке **Собственные значения**. Дисперсия, объясненная после-

довательностью собственных значений, предстанет в форме таблицы (рис. 2.137).

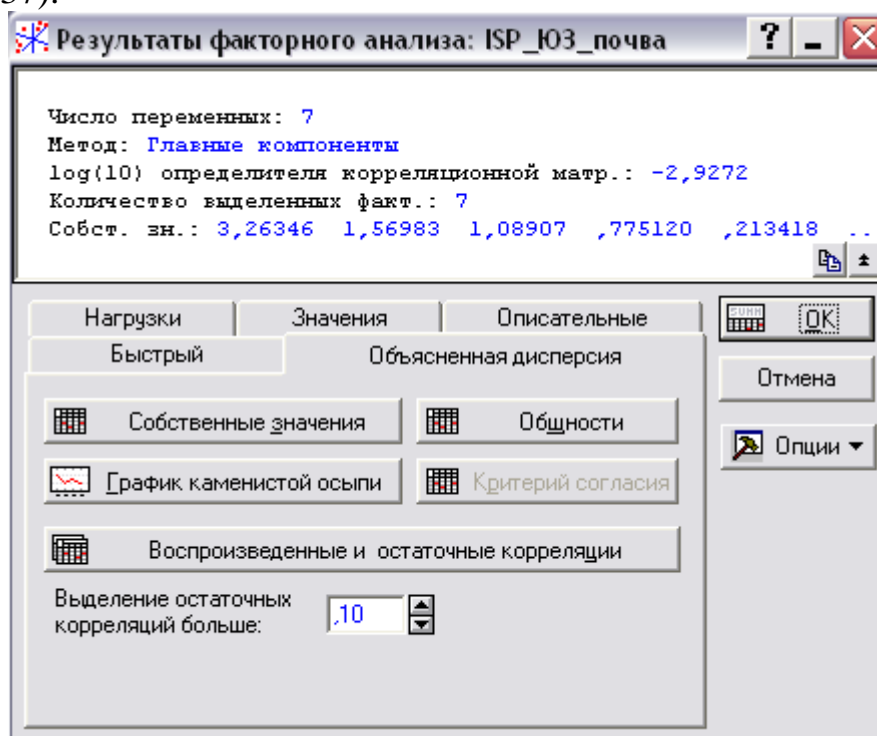


Рис. 2.136. Результаты факторного анализа для выборки ISP_Ю3_почва после применения метода главных компонент

Значен.	Собственные значения (ISP_Ю3_почва)			
	Соб. зн.	% общей дисперс.	Кумулятивн. собст. знач.	Кумулятивн. %
1	3,263460	46,62086	3,263460	46,6209
2	1,569828	22,42612	4,833288	69,0470
3	1,089074	15,55820	5,922362	84,6052
4	0,775120	11,07315	6,697482	95,6783
5	0,213418	3,04883	6,910900	98,7271
6	0,071073	1,01533	6,981973	99,7425
7	0,018027	0,25752	7,000000	100,0000

Рис. 2.137. Дисперсия, объясненная последовательностью всех собственных значений

В первом столбце (рис. 2.137) 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 $\approx 46,6\%$ от общей дисперсии, совокупность первых двух собственных значений – уже $\approx 69,0\%$, а совокупность первых трех собственных значений объясняет $\approx 84,6\%$ 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. 2.137, этот критерий предложил бы ограничиться первыми 3-мя собственными значениями.

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

Шаг 8. В режиме **Объяснимая дисперсия** (рис. 2.136) щелкните по кнопке **График каменистой осыпи** для просмотра графической информации о всех собственных значениях (рис. 2.138).

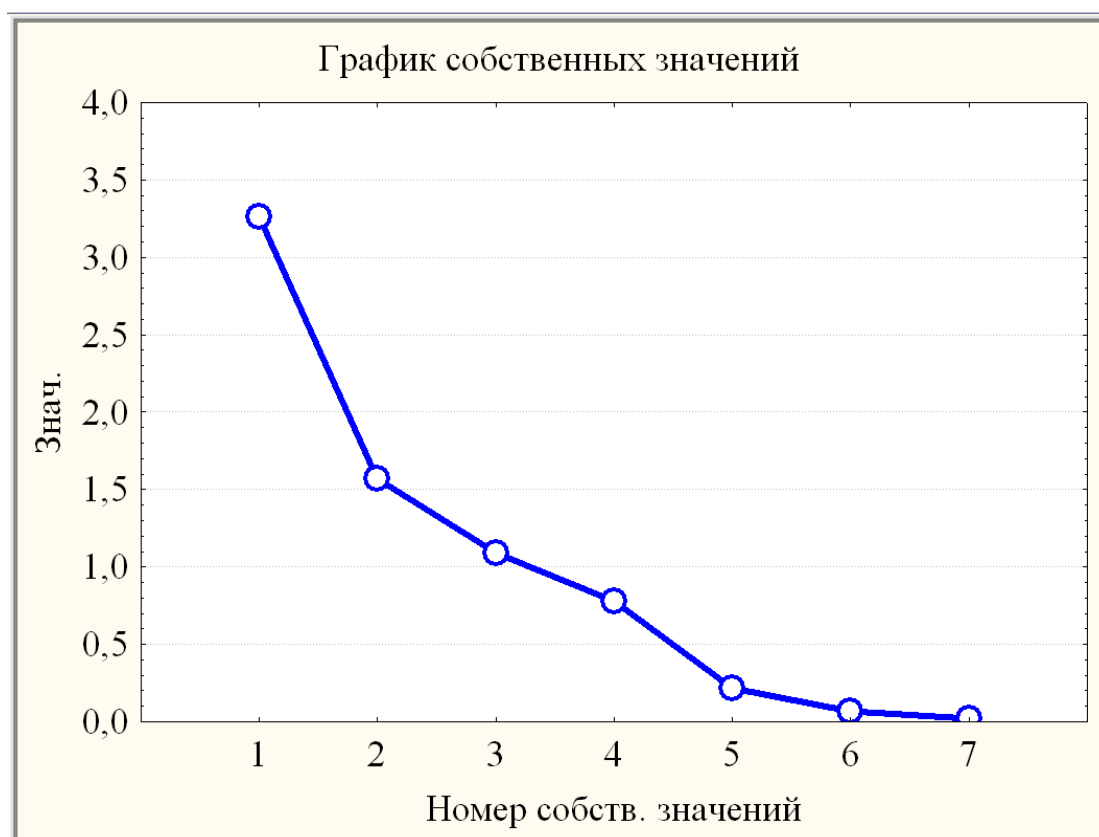


Рис. 2.138. График собственных значений

Рис. 2.138 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 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 9. Щелкните кнопкой **Отмена**, чтобы возвратиться в окно **Задайте метод выделения факторов** (рис. 2.135) в режиме **Дополнительно**, в котором в поле справа укажите Максимальное число факторов, равное 4 (в соответствии с результатами кластеризации микроэлементов на рис. 2.104) и щелкните **OK**. На экране появятся результаты анализа 4-х факторной модели (рис. 2.139), В режиме **Нагрузки** (рис. 2.139) выберите **Вращение факторов:** поворот варимаксимальный **Варимакс исходных** и щелкните по кнопке **Факторные нагрузки** для просмотра факторных нагрузок в выбранной повернутой 4-х факторной модели (рис. 2.140).

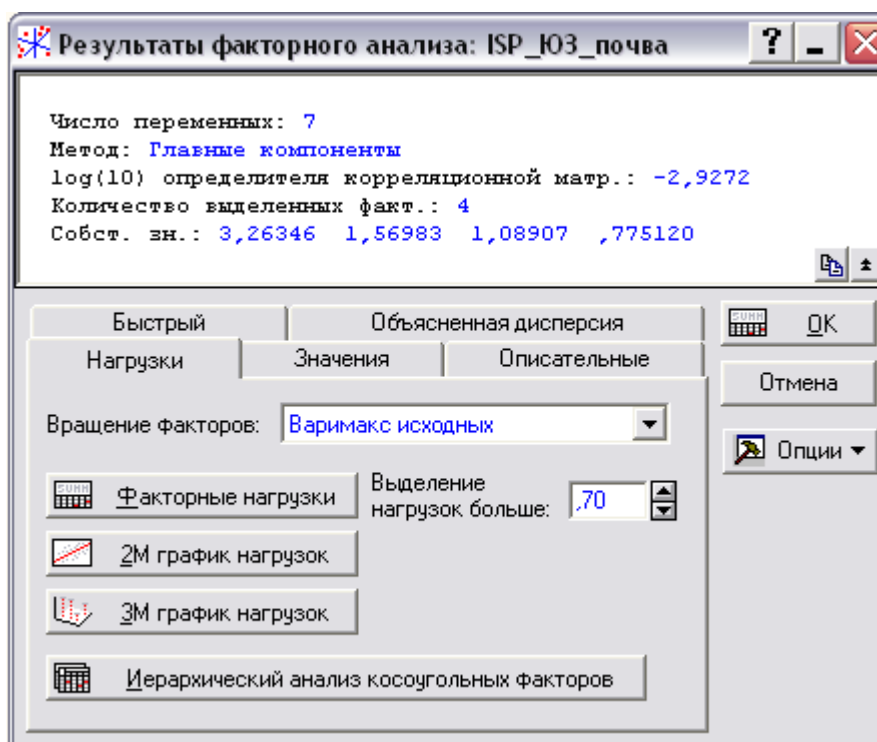


Рис. 2.139. Режим **Нагрузки** в окне **Результаты факторного анализа:**

Перемен.	Фактор. нагрузки (Варим. исх.) (ISP_ЮЗ_почва); Выделение: Главные компоненты (Отмечены нагрузки >,700000)			
	Фактор 1	Фактор 2	Фактор 3	Фактор 4
Na _{ISP}	0,845	0,453	-0,073	0,094
Fe _{ISP}	0,248	0,879	0,210	-0,311
Ba _{ISP}	0,127	-0,027	-0,092	0,974
Sc _{ISP}	0,172	0,951	-0,102	0,172
Cr _{ISP}	0,021	0,030	0,994	-0,083
Co _{ISP}	0,959	0,194	0,031	-0,073
La _{ISP}	0,873	0,032	0,080	0,333
Общ. дис.	2,504	1,922	1,064	1,207
Доля общ	0,358	0,275	0,152	0,172

Рис. 2.140. Вращаемые факторные нагрузки в выбранной 4-х факторной модели

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.

Для предания факторам простой интерпретации пространство факторов подвергнуто вращению так, что факторы отмечены высокими нагрузками для одних переменных, низкими нагрузками для других, и когда 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).

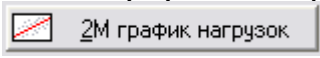
Согласно рис. 2.140, высокие факторные нагрузки микроэлементов распределились по факторам следующим образом:

Фактор Ф1 – наиболее весомый, составной, характеризуется в основном микроассоциацией Na_{ISP} – Co_{ISP} – La_{ISP}, связанной с Ф1 положительной корреляционной связью.

Фактор Ф2 – менее весомый, составной, характеризуется в основном микроассоциацией Fe_{ISP} – Sc_{ISP}, связанной с Ф2 положительной корреляционной связью.

Фактор Ф3 – характеризуется в основном микроэлементом Cr_{ISP}, связанным с Ф3 положительной корреляционной связью.

Фактор Ф4 –характеризуется в основном микроэлементом Ba_{ISP} , связанным с Ф4 положительной корреляционной связью.

Шаг 11. Для геометрической интерпретации факторной модели щелкните, например, по кнопке  и выберите факторы для графика: Ф1 и Ф2 (рис. 2.141).

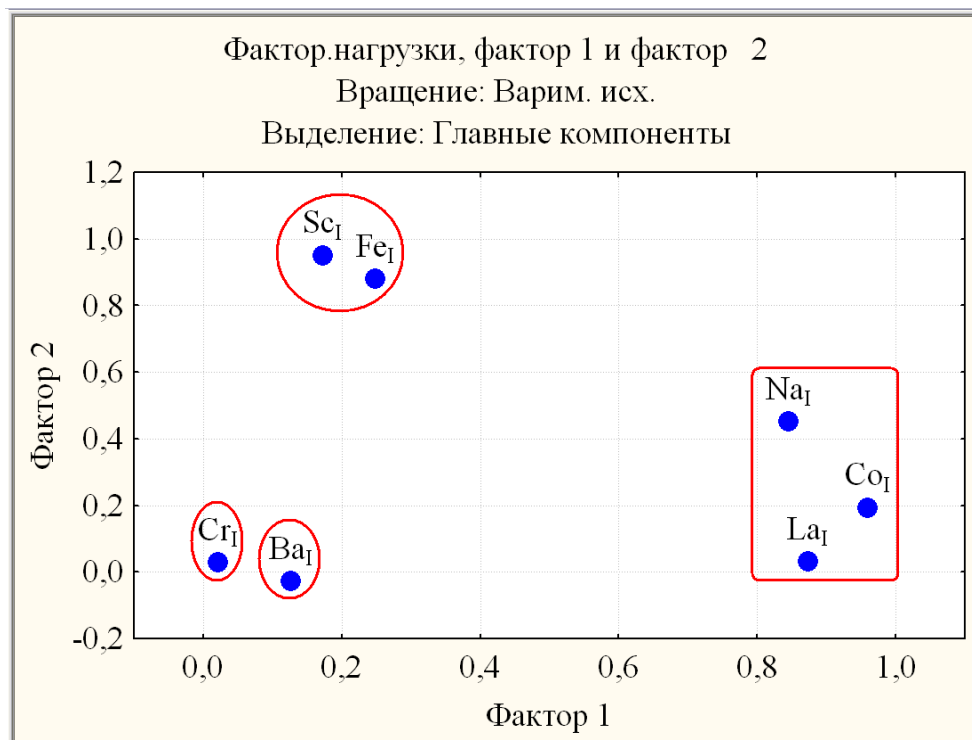
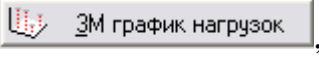

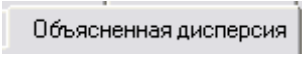
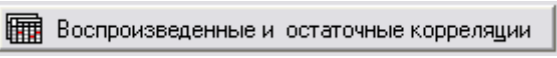


Рис. 2.141. Двумерный график факторных нагрузок микроэлементов выборки ISP_ЮЗ_почва

Диаграмма на рис. 2.141 наглядно иллюстрирует две микроассоциации $Na_{ISP} - Co_{ISP} - La_{ISP}$ и $Fe_{ISP} - Sc_{ISP}$, а также третью группу, содержащую два микроэлемента Cr_{ISP} и Ba_{ISP} , различия между которыми можно усмотреть с помощью , выбирая факторы для графика: {Ф1, Ф2, Ф3} или {Ф1, Ф2, Ф4}.

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

Шаг 12. В полученном ранее окне  в уже знакомом режиме  (рис.2.136) щелкните по кнопке  для просмотра воспроизведенной корреляционной матрицы (рис. 2.142)

Перемен.	Воспроизведенные корреляции (ISP_ЮЗ_почва)						
	Na _{ISP}	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}
Na _{ISP}	0,933	0,563	0,193	0,600	-0,049	0,889	0,778
Fe _{ISP}	0,563	0,976	-0,315	0,804	0,266	0,437	0,157
Ba _{ISP}	0,193	-0,315	0,974	0,173	-0,171	0,043	0,427
Sc _{ISP}	0,600	0,804	0,173	0,974	-0,084	0,334	0,230
Cr _{ISP}	-0,049	0,266	-0,171	-0,084	0,997	0,063	0,071
Co _{ISP}	0,889	0,437	0,043	0,334	0,063	0,963	0,822
La _{ISP}	0,778	0,157	0,427	0,230	0,071	0,822	0,881


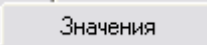

Рис. 2.142. Воспроизведенная корреляционная матрица выборки ISP_ЮЗ_почва

Сравнение исходной наблюдаемой (рис. 2.134) и воспроизведенной (рис. 2.142) корреляционных матриц выборки ISP_ЮЗ_почва выделяет в качестве самой контрастной пары Na_{ISP} – La_{ISP}, для которой значимость различия коэффициентов корреляции можно оценить аналогично рис.2.74:

Рис. 2.143. Сравнение исходного и воспроизведенного коэффициентов корреляции пары Na_{ISP} – La_{ISP}

Уровень значимости $p = 0,3242 > 0,10$ соответствует незначимому отличию. Таким образом, 4-х факторная модель обеспечивает воспроизведение исходных переменных с незначимыми отличиями. Заметим, что согласно рис.2.137 первые четыре собственных значений объясняют $\approx 95,7\%$ от общей дисперсии.

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

Шаг 13. В окне  Результаты факторного анализа войдите в режим  Значения (рис.2.144) и щелкните по кнопке  Сохранить значения для сохранения значений факторов по каждому наблюдению (рис. 2.145)

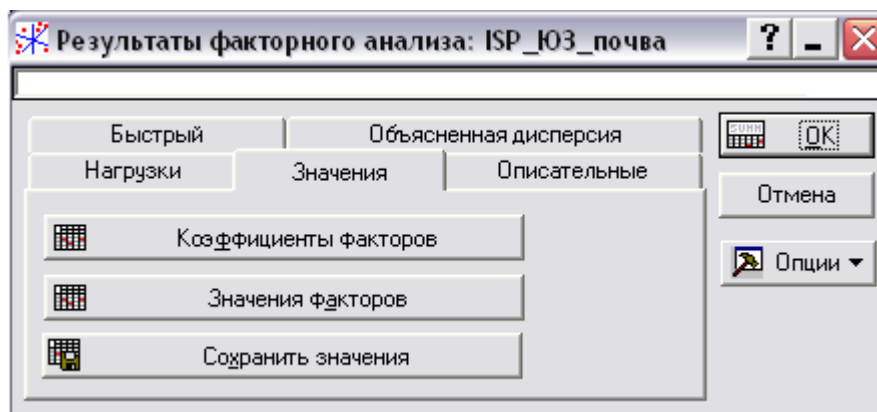


Рис. 2.144. Режим **Значения** в окне **Результаты факторного анализа:**

Данные: Таблица данных17* (11v * 22c)

	1	2	3	4	5	6	7	8	9	10	11
	Na _{ISP}	Fe _{ISP}	Ba _{ISP}	Sc _{ISP}	Cr _{ISP}	Co _{ISP}	La _{ISP}	Φ1	Φ2	Φ3	Φ4
1	2,16	3,16	370	10,2	144,5	21,1	28	0,126	-0,825	0,803	-0,264
2	2,27	3,22	350	9,1	161,2	21,6	26,2	0,278	-1,126	1,127	-1,068
3	2,32	3,61	330	10	127,7	23	24,9	0,325	-0,496	0,208	-1,910
4	2,54	3,82	380	13,4	136,7	22,6	33,1	0,671	0,706	0,648	0,300
5	2,46	3,26	320	9,7	136,4	22,1	27,5	0,702	-1,001	0,356	-1,945
6	1,99	2,78	420	10,2	213,5	16,8	26,2	-0,958	-0,818	2,663	1,742
7	2,47	2,99	380	11,2	73,5	20,9	27	0,205	-0,473	-1,120	-0,085
8	2,58	3,28	370	11,3	76,7	24	27,8	0,754	-0,340	-1,064	-0,609
9	2,62	3,62	350	11	173	22,8	26,4	0,464	-0,083	1,313	-0,997
10	2,23	2,36	430	9,8	103,5	17,2	27,6	-0,385	-1,327	-0,147	1,747
11	2,8	3,03	430	11,8	62,9	24,8	28,1	0,826	-0,220	-1,335	1,169
12	2,76	4	400	13,5	125	23,1	27,3	0,234	1,256	0,252	0,501
13	2,79	3,88	390	13,4	79,2	22,8	26,6	0,265	1,140	-0,966	0,067
14	2,92	4,18	400	14,3	148,4	24	31,1	0,728	1,519	0,875	0,718
15	2,21	2,54	410	10	64,9	19,9	30,8	0,332	-1,406	-1,090	0,976
16	2,45	2,58	390	10,9	64,8	19,8	29,7	0,380	-1,005	-1,294	0,447
17	2,96	3,72	390	13,6	128	23	28,9	0,611	0,983	0,205	0,383
18	2,6	3,85	380	14	115,4	23,3	31,4	0,586	1,009	0,031	0,181
19	2,78	3,97	380	14	116,3	23,1	30,2	0,577	1,219	0,001	0,090
20	1,52	2,97	380	10	100	11	18	-2,634	-0,309	-0,245	0,160
21	1,83	3,07	350	11	90	13	20	-1,920	-0,058	-0,688	-0,702
22	1,99	3,79	340	14	100	14	19	-2,169	1,654	-0,534	-0,901

Рис. 2.145. Значения старых (микроэлементов) и новых (факторов) переменных по каждому наблюдению для 4-х факторной модели выборки ISP_YO3_почва.

2.8. Кластерный анализ наблюдений в факторной модели ассоциации микроэлементов

Заключительной частью **этапа VI** является проведение кластерного анализа наблюдений в рамках построенной факторной модели содержаний микроэлементов, для чего необходимо выполнить последовательность следующих действий:

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

Шаг 2. Выберите **Многомерный разведочный анализ** в стартовой панели модуля и далее щелкните мышкой на разновидности **Кластерный анализ**. В появившемся окне выделите разновидность **Иерархическая классификация** и щелкните **OK** (рис. 2.146).

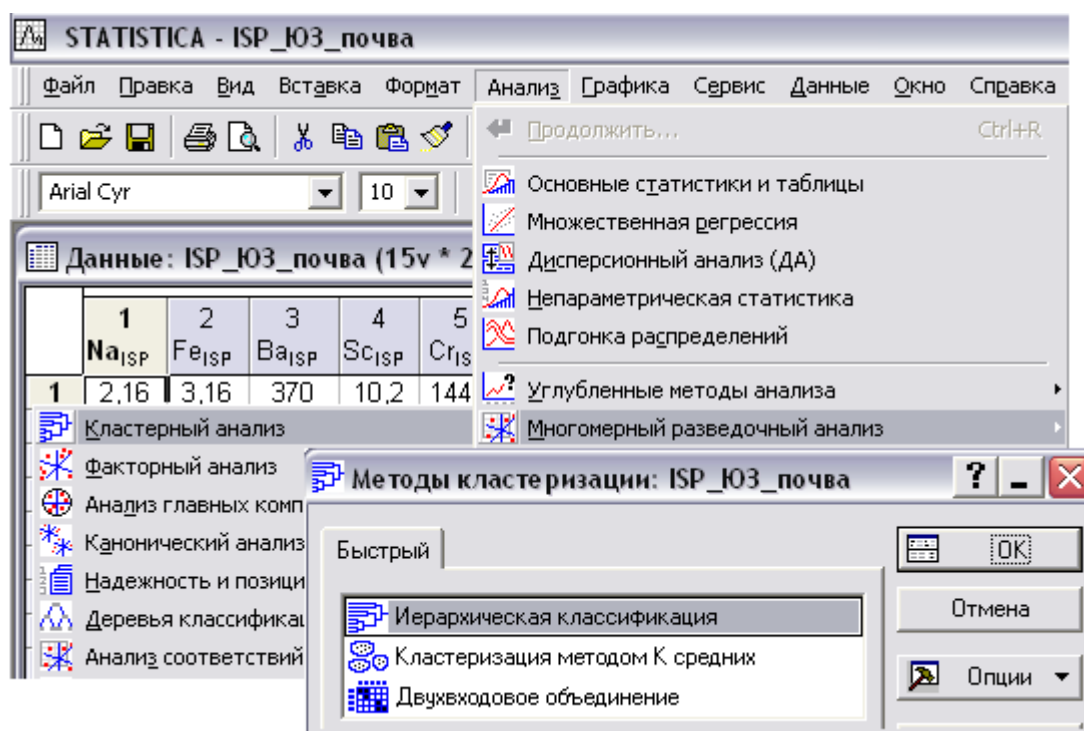


Рис. 2.146. Выбор **Иерархическая классификация** в модуле **Анализ**

Шаг 3. В открывшемся окне **Кластерный анализ** нажмите кнопку **Переменные:**, выберите все факторные переменные и нажмите **OK**. Далее в окне **Объекты:** выделите **Наблюдения (строки)**, в окне **Мера близости:** вы-

делите **Евклидово расстояние**, а в окне **Правило объединения:** – **Метод Варда** (рис. 2.147) и нажмите **OK**

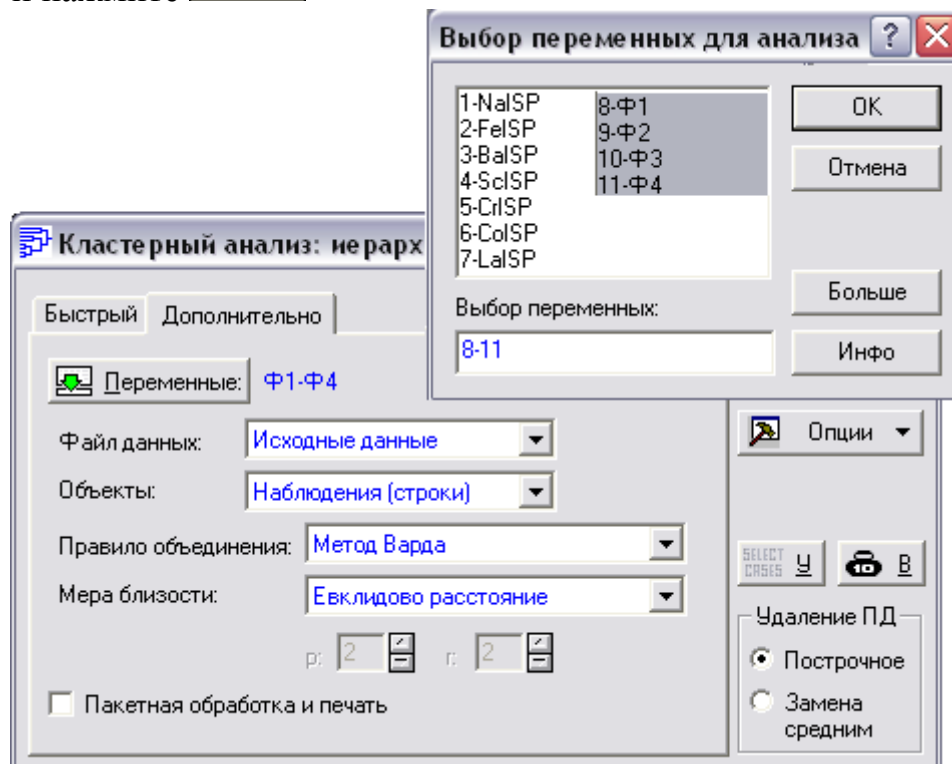


Рис. 2.147. Окно **Кластерный анализ:** для наблюдений:
Иерархическая классификация

Шаг 5. После этого щелкните **Вертикальная дендрограмма** (рис.2.148).

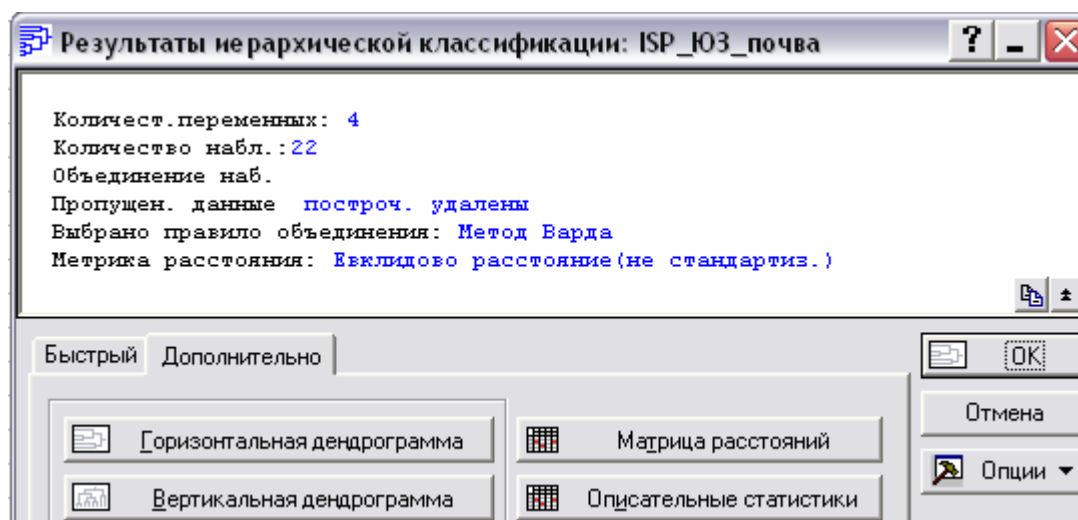


Рис. 2.148. Окно **Результаты иерархической классификации:**

Следующая дендрограмма наблюдений появится на экране:

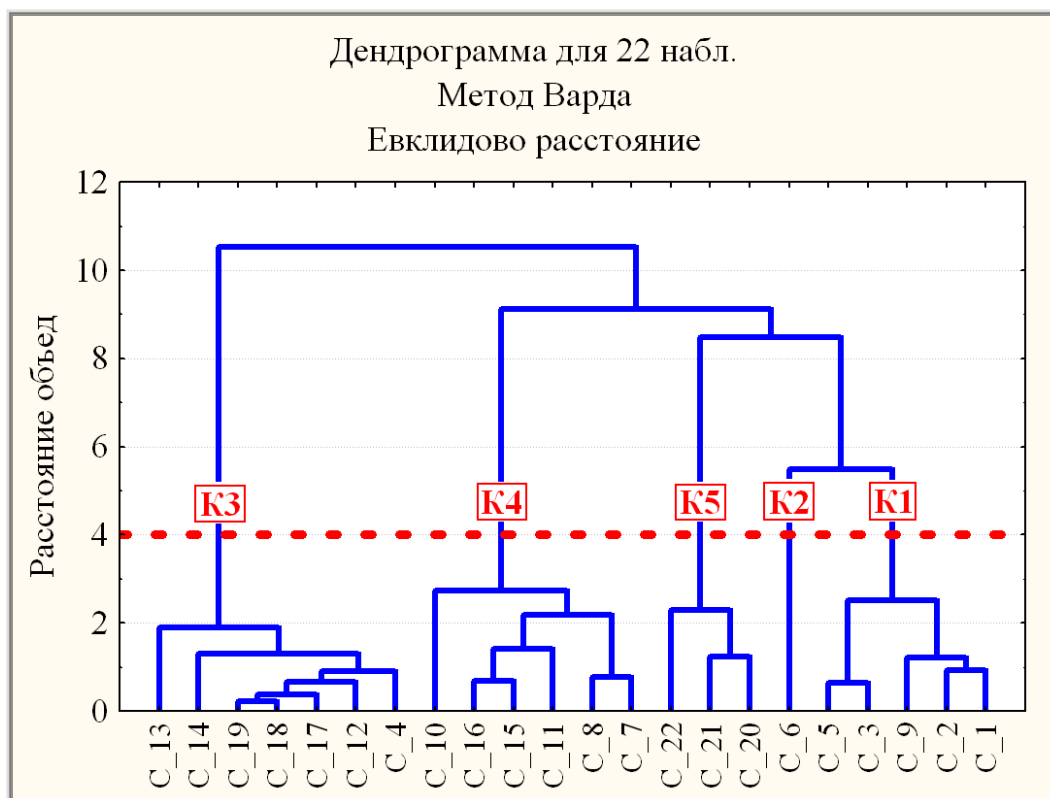









Рис. 2.149. Дендрограмма наблюдений в пространстве $\{\Phi_1, \Phi_2, \Phi_3, \Phi_4\}$

Таким образом, в построенном 4-х мерном факторном пространстве $\{\Phi_1, \Phi_2, \Phi_3, \Phi_4\}$ проведена кластеризация наблюдений. При этом в качестве меры близости выбрано евклидово расстояние, а в качестве правила объединения двух кластеров использован дисперсионный по сути метод Уорда (Варда). Методом древовидной кластеризации построено иерархическое дерево (рис. 2.149). В зависимости от выбора расстояния объединения можно получить соответствующее число кластеров. Так, например, расстоянию объединения, равного 4 (красная пунктирная горизонтальная прямая) соответствует 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 кластера-

ми, оценивает качество кластеризации наблюдений по каждому фактору посредством параметрического дисперсионного анализа.

Шаг 6. В окне  **Методы кластеризации:** (рис. 2.124) выберите метод  **Кластеризация методом К средних** и щелкните  **OK**.

Шаг 7. В открывшемся окне  **Кластерный анализ:** нажмите кнопку  **Переменные:**, выберите все факторные переменные аналогично рис. 2.147 и нажмите  **OK**. В окне **Объекты:** выделите **Наблюдения (строки)**, в окне **Число кластеров:** выделите 5 в соответствии с рис. 2.149, а в окне **Начальные центры кластеров** — **Сортировать расстояния** (рис. 2.150) и нажмите  **OK** для просмотра результатов кластеризации наблюдений методом *K*-средних (рис. 2.151)

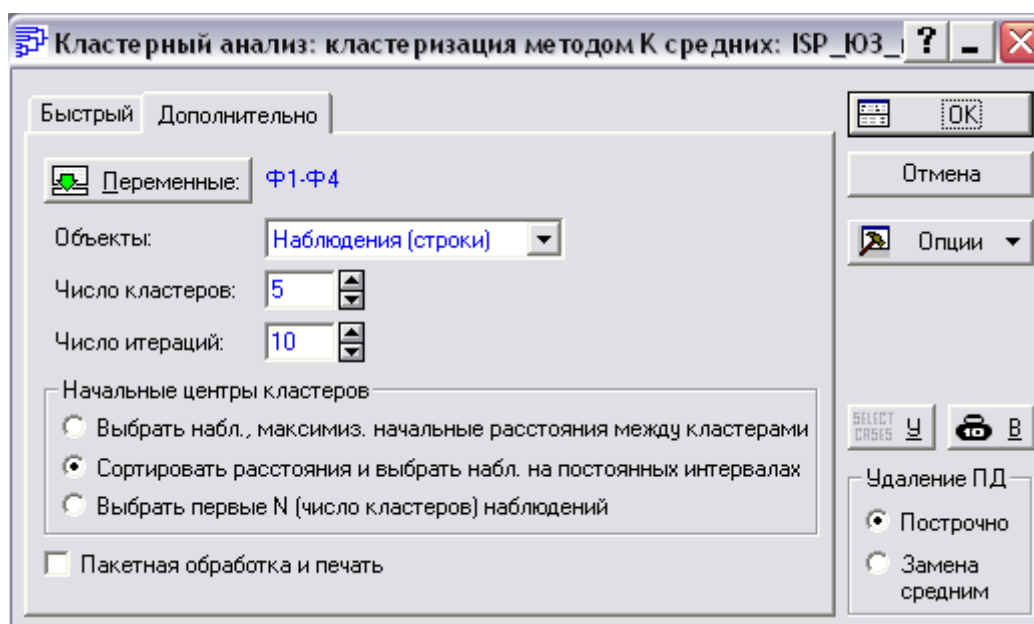



Рис. 2.150. Окно  **Кластерный анализ:** для наблюдений:



Заметим, что поле ввода «Число итераций» позволяет пользователю указать максимальное число итераций. Метод *K*-средних является итеративной процедурой; на каждой итерации объекты перемещаются в различные кластеры. Алгоритм, реализованный в  **Кластерный анализ:**, весьма эффективен и обычно нет необходимости менять установку по умолчанию (10 итераций).

Окно **Начальные центры кластеров** содержит три опции, позволяющие выбрать способ вычисления начального расположения центров кластеризации. Заметим, что результаты метода К-средних до некоторой степени зависят от начальной конфигурации (т.е. расположения кластерных средних или центров). Это, в частности, происходит, когда имеется много небольших кластеров, которые ясно различимы. Опция **«Максимизация расстояний между кластерами»** выбирает перебором наблюдения, которые служат центрами кластеров, 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.

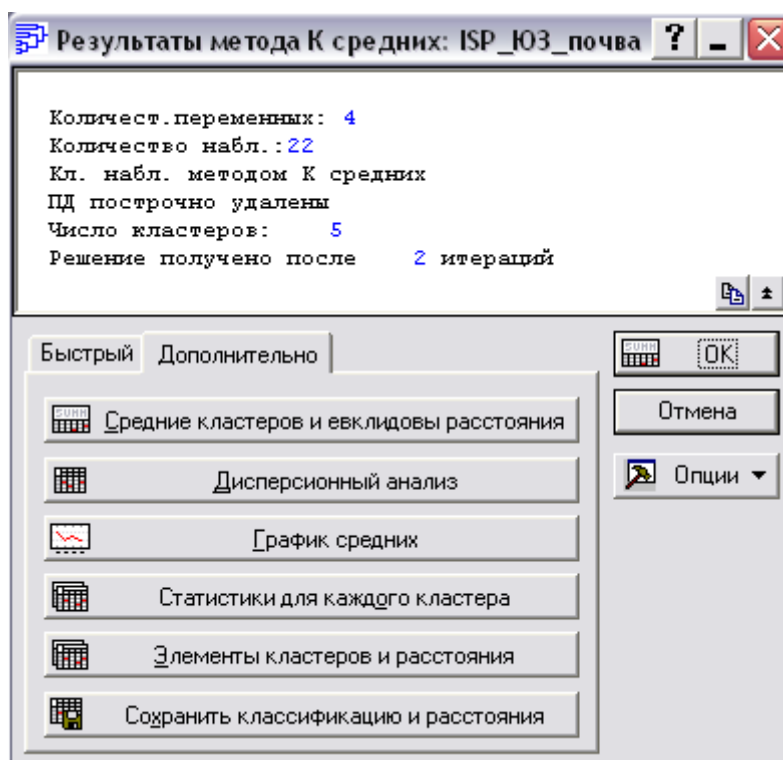



Рис. 2.151. Окно **Результаты метода К средних:**


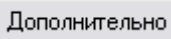
Рассмотрим результаты кластеризации наблюдений методом K -средних (рис. 2.129)

Шаг 8. Нажмите прежде всего кнопку  Дисперсионный анализ для построения стандартной таблицы результатов дисперсионного анализа кластеризации наблюдений по каждому фактору (рис. 2.152).

перемен.	Дисперсионный анализ (ISP_ЮЗ_почва)					
	Между SS	сс	Внутри SS	сс	F	значим. p
Ф1	19,369	4	1,632	17	50,454	0,000000
Ф2	16,273	4	4,727	17	14,632	0,000024
Ф3	16,966	4	4,034	17	17,876	0,000006
Ф4	14,310	4	6,690	17	9,091	0,000407

Рис. 2.152. Результаты дисперсионного анализа кластеризации наблюдений по факторам

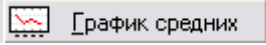
Для оценки качества классификации можно сравнить изменчивость внутри кластера (мала, если классификация хорошая) с изменчивостью вне кластера (велика, если классификация хорошая), что осуществляется путём стандартного межгруппового дисперсионного анализа для каждого фактора. Значения F -статистики, полученные для каждого фактора, является индикатором того, насколько хорошо соответствующий фактор дискриминирует кластеры. Значения p -уровней определяют меры статистической значимости (на уровне значимости $p < 0,0005$) различий между кластерами. Желательно получить сильно различающиеся средние для большинства факторов, используемых в анализе.

Шаг 9. 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. 2.153).

перемен.	Средн.класт. (ISP_ЮЗ_почва)				
	Кластер No. 1	Кластер No. 2	Кластер No. 3	Кластер No. 4	Кластер No. 5
Ф1	0,379	-0,958	0,525	0,352	-2,241
Ф2	-0,706	-0,818	1,119	-0,795	0,429
Ф3	0,762	2,663	0,149	-1,008	-0,489
Ф4	-1,237	1,742	0,320	0,608	-0,481

Рис. 2.153. 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. 2.154).

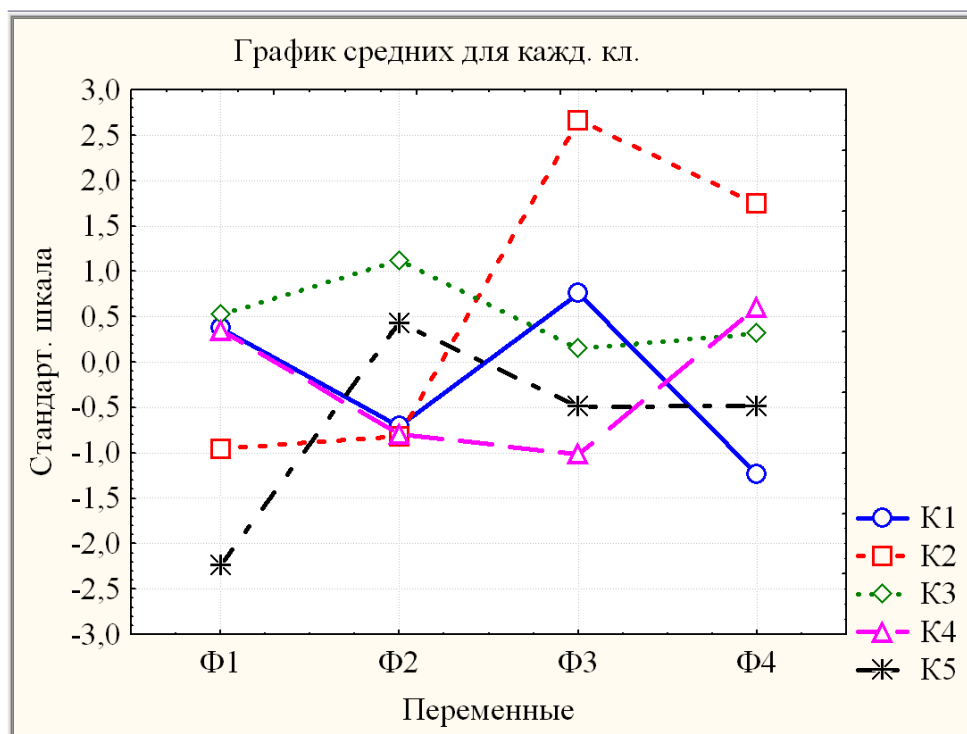




Рис. 2.154. Линейные графики факторных средних для каждого кластера

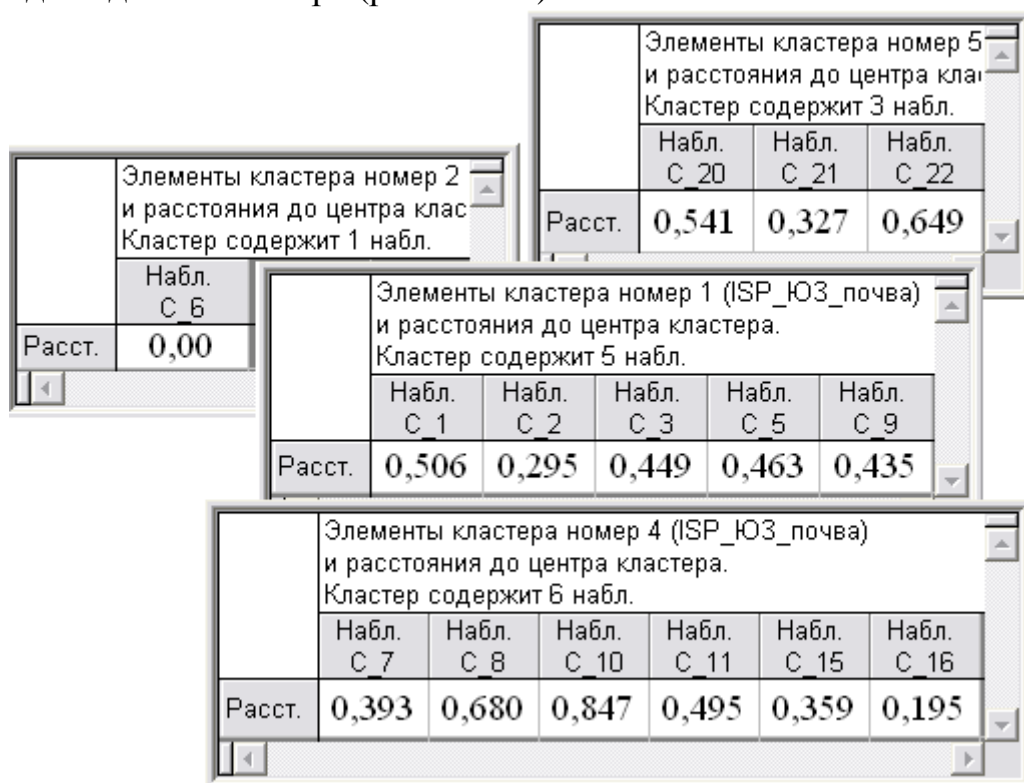
Этот график полезен для визуального обобщения различий средних между кластерами.

Шаг 10. Нажмите  для вывода на экран списка элементов (наблюдений) каждого кластера, где для каждого наблюдения вычислено его расстояние до центра кластера (рис. 2.155).

Элементы кластера номер 3 (ISP_Ю3_почва) и расстояния до центра кластера. Кластер содержит 7 набл.							
	Набл. С 4	Набл. С 12	Набл. С 13	Набл. С 14	Набл. С 17	Набл. С 18	Набл. С 19
Расст.	0,332	0,191	0,586	0,471	0,091	0,111	0,148

Рис. 2.155. Элементы кластера К3 с расстояниями до центра

Опция  **Элементы кластеров и расстояния** выводит на экран евклидовы расстояния объектов (наблюдений) от центров (средних значений) соответствующих им кластеров. Это позволит идентифицировать потенциально "плохие" элементы кластеров, т.е. такие объекты, слишком далеко расположенные от центра кластера, но, по-видимому, не принадлежащие к какому-либо другому кластеру (они ещё дальше расположены от центров других кластеров). Эти расстояния приведены в стандартных таблицах результатов, каждая таблица результатов показывает результаты для одного кластера (рис. 2.155').



Элементы кластера номер 2 и расстояния до центра кластера. Кластер содержит 1 набл.	
Набл.	С 6
Расст.	0,00


Элементы кластера номер 5 и расстояния до центра кластера. Кластер содержит 3 набл.	
Набл.	С 20
Набл.	С 21
Набл.	С 22
Расст.	0,541
Расст.	0,327
Расст.	0,649

Элементы кластера номер 1 (ISP_Ю3_почва) и расстояния до центра кластера. Кластер содержит 5 набл.					
Набл.	С 1	Набл.	С 2	Набл.	С 3
Набл.	С 5	Набл.	С 9		
Расст.	0,506	0,295	0,449	0,463	0,435

Элементы кластера номер 4 (ISP_Ю3_почва) и расстояния до центра кластера. Кластер содержит 6 набл.					
Набл.	С 7	Набл.	С 8	Набл.	С 10
Набл.	С 11	Набл.	С 15	Набл.	С 16
Расст.	0,393	0,680	0,847	0,495	0,359
Расст.				0,195	

Рис. 2.155'. Элементы кластеров K1, K2, K4, K5 с расстояниями

Полученные результаты кластеризации наблюдений методом *K*-средних (с опцией «Сортировать расстояния и взять наблюдения на постоянных интервалах»,) в полной мере соответствуют иерархическому дереву (рис. 2.149), 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  **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. 2.155").

ISP_ЮЗ_почва							
	1 Ф1	2 Ф2	3 Ф3	4 Ф4	5 НАБЛ.	6 КЛАСТЕР	7 РАССТ.
C_1	0,126	-0,825	0,803	-0,264	1	1	0,51
C_2	0,278	-1,126	1,127	-1,068	2	1	0,30
C_3	0,325	-0,496	0,208	-1,910	3	1	0,45
C_4	0,671	0,706	0,648	0,300	4	3	0,33
C_5	0,702	-1,001	0,356	-1,945	5	1	0,46
C_6	-0,958	-0,818	2,663	1,742	6	2	0,00
C_7	0,205	-0,473	-1,120	-0,085	7	4	0,39
C_8	0,754	-0,340	-1,064	-0,609	8	4	0,68
C_9	0,464	-0,083	1,313	-0,997	9	1	0,43
C_10	-0,385	-1,327	-0,147	1,747	10	4	0,85
C_11	0,826	-0,220	-1,335	1,169	11	4	0,49
C_12	0,234	1,256	0,252	0,501	12	3	0,19
C_13	0,265	1,140	-0,966	0,067	13	3	0,59
C_14	0,728	1,519	0,875	0,718	14	3	0,47
C_15	0,332	-1,406	-1,090	0,976	15	4	0,36
C_16	0,380	-1,005	-1,294	0,447	16	4	0,19
C_17	0,611	0,983	0,205	0,383	17	3	0,09
C_18	0,586	1,009	0,031	0,181	18	3	0,11
C_19	0,577	1,219	0,001	0,090	19	3	0,15
C_20	-2,634	-0,309	-0,245	0,160	20	5	0,54
C_21	-1,920	-0,058	-0,688	-0,702	21	5	0,33
C_22	-2,169	1,654	-0,534	-0,901	22	5	0,65

Рис. 2.155". Результаты факторной классификации наблюдений

Для каждого фактора (рис.2.154) согласно апостериорному критерию аналогично рис. 2.42-43 можно выделить однородные (различающиеся незначимо, то есть на уровне значимости $p > 0,10$) иногда пересекающиеся группы кластеров. Предварительно оформим стандартную процедуру дисперсионного анализа результатов факторной классификации наблюдений.

Шаг 11. Запустите **Анализ** и щелкните **Дисперсионный анализ (ДА)** (см. рис. 2.34). В появившемся окне выберите тип анализа (однофакторный дисперсионный анализ) и задайте метод (диалог быстрых спецификаций), как показано на рис.2.35, и нажмите **ОК**. В появившемся окне выберите в качестве зависимых все факторные переменные (позиции 8-11), а в качестве категориальной **КЛАСТЕР** (позиция 12), закажите все коды (1-5), как показано на рис.2.156, и нажмите **ОК**.

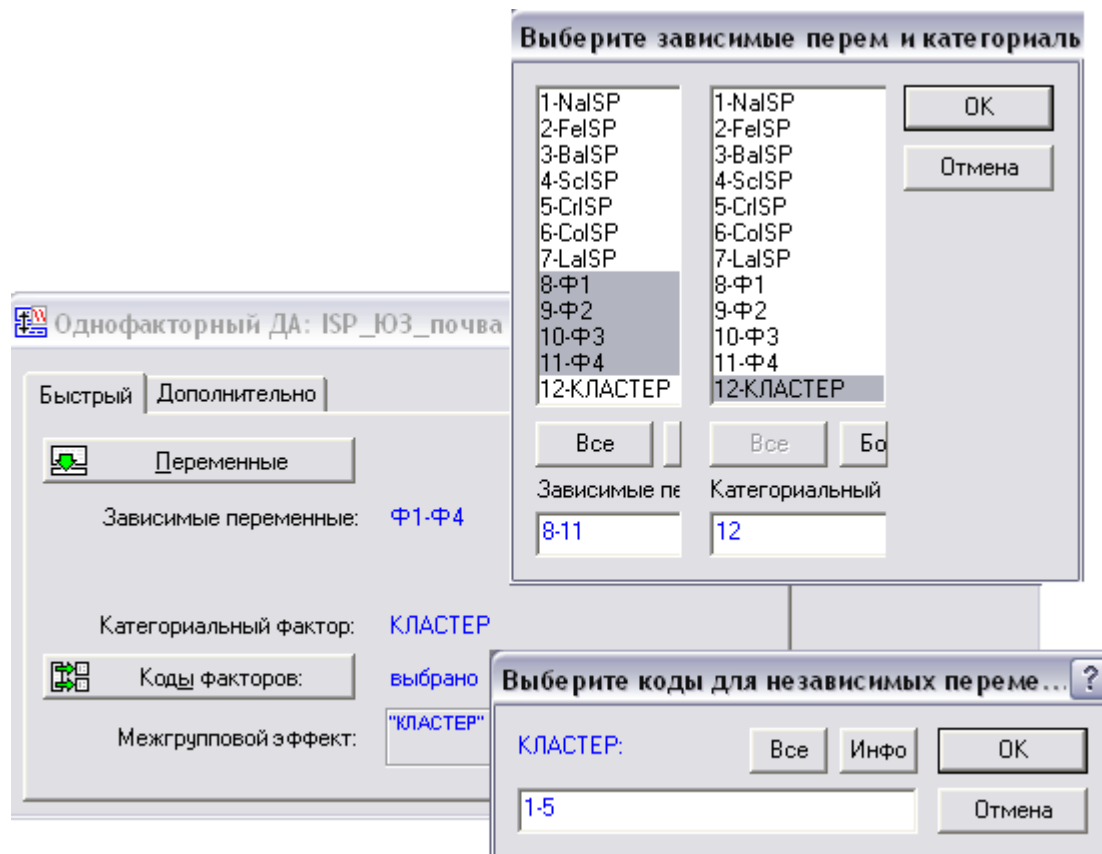


Рис. 2.156. Выбор переменных для дисперсионного кластеров

Шаг 12. После этого в появившемся окне **Результаты анализа** (рис.2.37) в режиме **Итоги** нажмите **Статистики ячеек** для просмотра результатов описательной статистики (расширенная таблица средних на рис.2.153) и **Одномерн. результаты** для просмотра результатов дисперсионного анализа (рис.2.152).

Продemonстрируем возможности модуля **Дисперсионный анализ (ДА)** для расширенного анализа полученного результата. Так как значения одномерного F -критерия для каждого фактора – значимы (рис.2.152), то

далее можно использовать метод множественных сравнений для выявления тех групп кластеров, которые вызвали данный эффект и максимально отличаются от других.

Шаг 13. В окне **Результаты анализа** (рис. 2.37) нажмите кнопку **Больше** и в открывшемся окне выберите режим **Апостериорные**, в котором представлены различные апостериорные критерии (рис. 2.157). Выберите, например, критерий **Фишера НЗР** с опцией отображения **Однородные группы:** или **Значимые различия** (рис. 2.136).

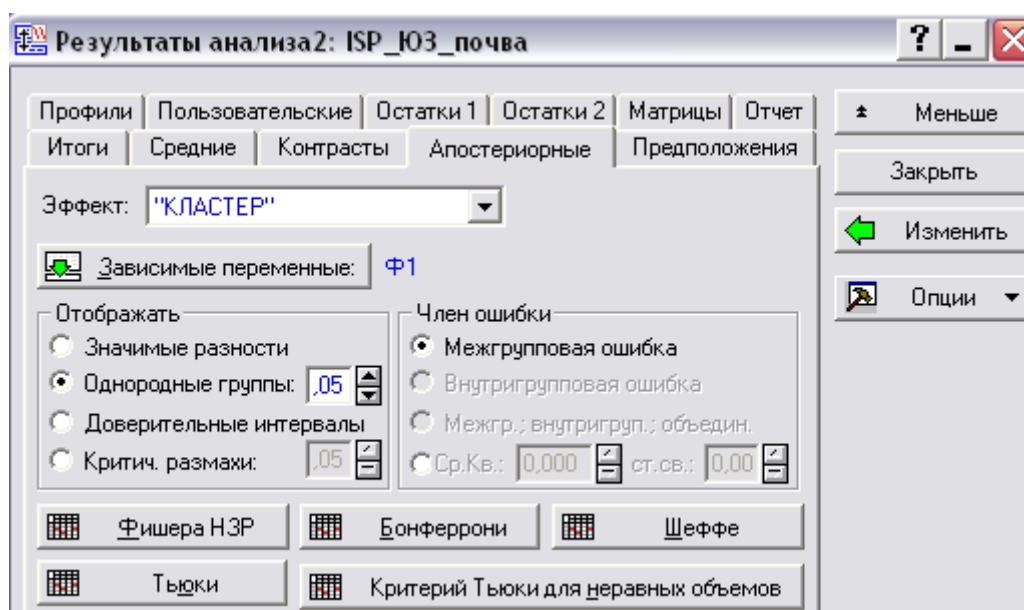


Рис. 2.157. Режим **Апостериорные** для выбора апостериорных критериев множественных сравнений средних

НЗР крит.; перем. Ф1 (ISP_Ю3_почва)
Однородные группы, alpha = ,05000
Ошибка: Межгр. MS = ,09597, сс = 17,

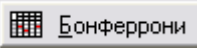

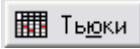
КЛАСТЕР	Ф1	1	2	3
	Среднее			
5	-2,241		****	
2	-0,958			****
4	0,352	****		
1	0,379	****		
3	0,525	****		

НЗР крит.; перем. Ф1 (ISP_Ю3_почва)
Вероятности для апостер. критериев
Ошибка: Межгр. MS = ,09597, сс = 17,0

КЛ	{1}	{2}	{3}	{4}	{5}
	0,38	-0,96	0,52	0,35	-2,24
1		0,001	0,432	0,889	0,000
2	0,001		0,000	0,001	0,002
3	0,432	0,000		0,331	0,000
4	0,889	0,001	0,331		0,000
5	0,000	0,002	0,000	0,000	

Рис. 2.158. Результаты сравнений кластерных средних для Ф1 по **Фишера НЗР**: **Однородные группы:** (слева) и **Значимые различия** (справа)

Согласно рис. 2.158, для фактора Φ_1 по апостериорному критерию наименьших значений разности (НЗР) выделены три однородные (различающиеся незначимо) группы кластеров, расположенные в порядке убывания средних: {К3, К1, К4}, {К2}, {К5}. Разделение на три группы кластеров является сильно значимым, так как К2 отличается как от К4, так и от К5 сильно значимо ($0,005 > p > 0,0005$).

Остальные апостериорные критерии , ,  подтверждают этот результат в несколько смягченной форме: К2 отличается как от К4, так и от К5 статистически значимо ($0,05 > p > 0,005$). Таким образом, критерий НЗР дает самые значимые результаты по сравнению с другими апостериорными критериями.

Аналогичным образом (на основании рис. 2.158 для Φ_1) можно оформить результаты сравнений кластерных средних для остальных факторов. Таким образом, :

Φ_1 : {К3, К1, К4}, {К2}, {К5}. Разделение на три группы кластеров является сильно значимым, так как К2 отличается как от К4, так и от К5 сильно значимо ($0,005 > p > 0,0005$).

Φ_2 : {К3, К5}, {К5, К2}, {К1, К4, К2}. Разделение на три пересекающиеся группы кластеров так, что К5 отличается как от К3 слабо значимо ($0,10 > p > 0,05$), а от К1 и К4 статистически значимо ($0,05 > p > 0,005$).

Φ_3 : {К2}, {К1, К3}, {К3, К5}, {К5, К4}. Разделение на 4-е группы кластеров, из которых три последовательно пересекающиеся. При этом К2 и К1, К1 и К5, К3 и К4 различаются статистически значимо ($0,05 > p > 0,005$).

Φ_4 : {К2, К4, К3}, {К4, К3, К5}, {К5, К1}. Разделение на три пересекающиеся группы кластеров так, что К5 отличается как от К2 статистически значимо ($0,05 > p > 0,005$), а от К1 и К3 сильно значимо ($0,005 > p > 0,0005$).

Результаты кластерного анализа наблюдений по совокупности факторов (рис. 2.153-154) с учетом результатов сравнений кластерных средних для каждого фактора позволяют провести классификацию наблюдений в порядковой шкале стандартизированных измерений, полагая в качестве уровня «Выше среднего» - стандартизированный интервал (0; +1,5) наблюдений, «Ниже среднего» – (-1,5; 0). Аномально высокие значения ($> +1,5$) определяют уровень «**Лидер**», а аномально низкие значения ($< -1,5$) определяют уровень «**Аутсайдер**».

Таблица 2.10

Классификация наблюдений по совокупности факторов в порядковой шкале стандартизированных измерений.

Кластер	Уровень кластера на фоне среднего по фактору				Состав кластера
	Ф1 (Na – Co – La)	Ф2 (Fe – Sc)	Ф3 (Cr)	Ф4 (Ba)	
К1	Выше среднего	Ниже среднего	Выше среднего	Ниже среднего	C_1+2+3+5+9
К2	Ниже среднего	Ниже среднего	Лидер	Лидер	C_6
К3	Выше среднего	Выше среднего	Выше среднего	Выше среднего	C_4+12+13+14+ +17+18+19
К4	Выше среднего	Ниже среднего	Ниже среднего	Выше среднего	C_7+8+10+11+ +15+16
К5	Аут- сайдер	Выше среднего	Ниже среднего	Ниже среднего	C_20+21+22

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 теста Краскела-Уоллиса.

Шаг 14. Запустите модуль **Анализ** и щелкните в нем по кнопке **Непараметрическая статистика** (см. рис.2.34). Затем выберите сравнение нескольких независимых групп **Сравнение нескольких независимых групп** (рис.2.47), закажите переменные стандартным образом (см. рис. 2.156) и щелкните по кнопке **Да Краскела-Уоллиса и медианный тест** (рис.2.159). В итоге на экране появятся результаты непараметрического однофакторного дисперсионного анализа по тесту Краскела-Уоллиса (рис.2.160).

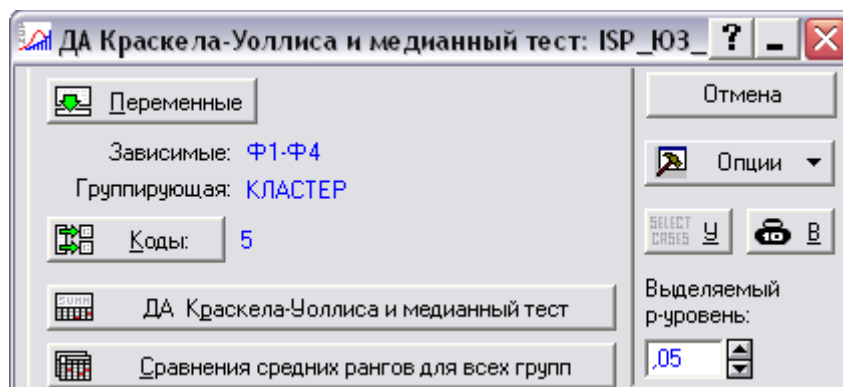


Рис.2.159. Окно непараметрического однофакторного ДА

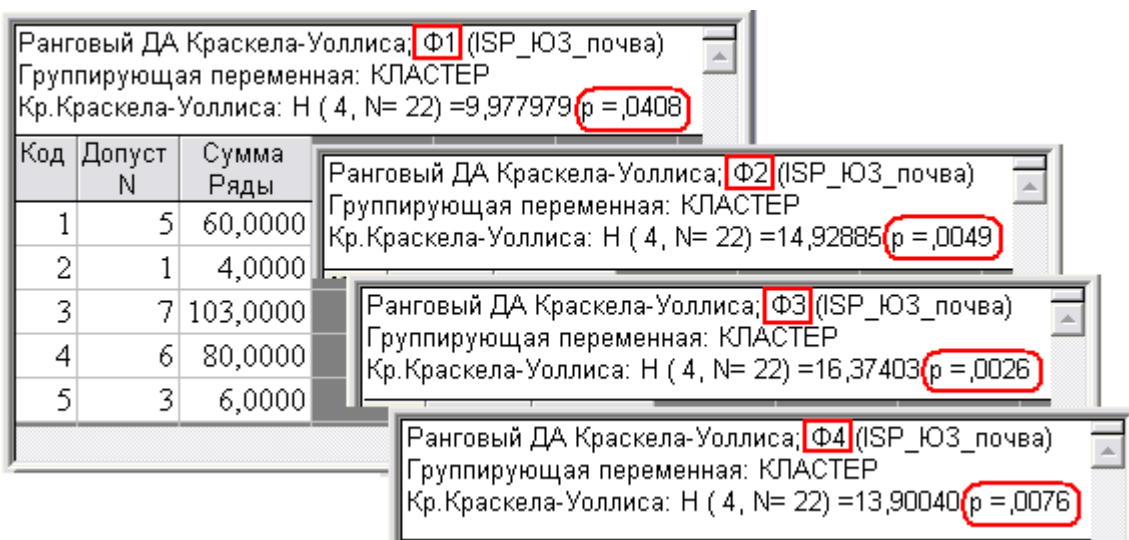
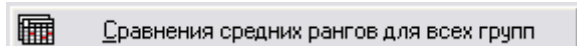


Рис.2.160. Результаты рангового дисперсионного анализа кластеризации по тесту Краскела-Уоллиса

Согласно критерию Краскела-Уоллиса (рис. 2.160) совокупность кластеров К1-5 различается статистически значимо в случае Ф1 ($0,050 > p \approx 0,041 > 0,005$) и Ф4 ($0,050 > p \approx 0,008 > 0,005$), а в случае Ф2 ($0,0050 > p \approx 0,0049 > 0,0005$) и Ф3 ($0,0050 > p \approx 0,0026 > 0,0005$) - сильно значимо.

Так как для каждого фактора существуют значимые различия по совокупности кластеров К1-5, то далее можно использовать непараметрический аналог метода множественных сравнений.

Шаг 15. В окне **ДА Краскела-Уоллиса** (рис. 2.159) нажмите кнопку



(рис. 2.161).

Зависим. Ф1	Сравнения p значений (2-стороннее); Ф1 (ISP_Ю3_почва)				
	1	2	3	4	5
	R:12,000	R:4,0000	R:14,714	R:13,333	R:2,0000
1		1,000	1,000	1,000	0,350
2	1,000		1,000	1,000	1,000
3	1,000	1,000		1,000	0,045
4	1,000	1,000	1,000		0,136
5	0,350	1,000	0,045	0,136	

Рис. 2.161. Результаты сравнений средних рангов кластеров для Ф1

Согласно рис. 2. 161, в случае Ф1 совокупность кластеров К1-К5 различается статистически значимо исключительно за счет статистически значимого различия между кластерами К5 и К3 ($0,050 > p \approx 0,045 >$

0,005). При этом моно кластер К2, состоящий из одного наблюдения С_6, оценивается как незначимо отличающийся от остальных.

Согласно рис. 2.161, для фактора Ф1 по ранговому критерию выделены две однородные (различающиеся незначимо) пересекающиеся группы кластеров, расположенные в порядке убывания средних рангов: {К3, К4, К1, К2} и {К4, К1, К2, К5} так, что К3 отличается от К5 статистически значимо ($0,050 > p \approx 0,045 > 0,005$).

В силу расхождения результатов непараметрического и параметрического однофакторного дисперсионного анализа в связи с малочисленностью состава кластеров более корректными признаются результаты непараметрического дисперсионного анализа.

Аналогичные (рис. 2.161 для Ф1) результаты сравнений средних рангов кластеров для остальных факторов представлены на рис. 2.161'.

Сравнения p значений (2-стороннее); Ф2 (ISP_Ю3_почва) Группирующая переменная: КЛАСТЕР Кр.Краскела-Уоллиса: H (4, N= 22) =14,92885 p = ,0049				
1	2	3	4	5
R:7,0000	R:7,0000	R:18,000	R:6,3333	R:15,667
	1,000	0,038	1,000	0,676
1,000		1,000	1,000	1,000
0,038	1,000		0,012	1,000
				0,421

Сравнения p значений (2-стороннее); Ф3 (ISP_Ю3_почва) Группирующая переменная: КЛАСТЕР Кр.Краскела-Уоллиса: H (4, N= 22) =16,37403 p = ,0026				
1	2	3	4	5
R:17,800	R:22,000	R:13,286	R:4,1667	R:8,0000
	1,000	1,000	0,005	0,388
1,000		1,000	0,110	0,619
1,000	1,000		0,116	1,000
				1,000

Сравнения p значений (2-стороннее); Ф4 (ISP_Ю3_почва) Группирующая переменная: КЛАСТЕР Кр.Краскела-Уоллиса: H (4, N= 22) =13,90040 p = ,0076				
1	2	3	4	5
R:3,6000	R:21,000	R:14,000	R:15,500	R:7,6667
	0,144	0,062	0,025	1,000
0,144		1,000	1,000	0,754
0,062	1,000		1,000	1,000
0,025	1,000	1,000		0,880
1,000	0,754	1,000	0,880	

Рис. 2.161'. Результаты сравнений средних рангов кластеров для Ф2-Ф4

Согласно рис. 2.161', в случае Ф2 кластер К3 отличается статистически значимо как от К1 ($0,050 > p \approx 0,038 > 0,005$) так и от К4 ($0,050 > p \approx 0,012 > 0,005$). В случае Ф3 кластер К4 отличается статистически значимо от К1 ($0,050 > p \approx 0,0053 > 0,005$) и близок к слабо значимому отличию от К2 и К3 ($p \approx 0,10$). В случае Ф4 кластер К1 отличается статистически значимо от К4 ($0,050 > p \approx 0,025 > 0,005$) и слабо значимо от К3 ($0,10 > p \approx 0,06 > 0,05$). Различие между К1 и К2, близкое к слабо значимому, в значительной степени объясняется моно природой К2.

С учетом ослабленного влияния моно кластера К2 на результаты сравнений средних рангов кластеров для каждого фактора, можно упростить классификацию наблюдений по совокупности факторов (табл.2.10), упразднив уровень «Лидер», порожденный результатами кластера К2.

Эколого-геохимическая интерпретация результатов кластеризации по ассоциации микроэлементов (I Микроэлементы) и территориальному расположению (II Наблюдения) позволяют сделать следующие выводы:

I) Микроэлементы -

1) {Fe – Sc – La; Co – Cr} по методу ИНАА и {Fe – Sc – La – Na; Co – Cr – Ba} по методу ISP, рис. 2.99-100. Результаты анализов различными методами (ИНАА – инструментальный нейтронно-активационный анализ, ISP – атомно-эмиссионный анализ с индуктивно связанной плазмой) позволяют выявить близкую ассоциацию элементов. В данном случае это характеризует, во-первых, хорошую сходимость результатов анализов и во-вторых, специфику производства, которое ориентировано на геохимическую специализацию.

2) {NaISP – CoISP – LaISP, FeISP – ScISP, BaISP, CrISP}, рис. 2.204 и, как следствие, интерпретация факторов, рис. 2.118-119. В данном случае фиксируется хорошо выраженная связь элементов, которая объясняется близкими геохимическими особенностями и свойствами..

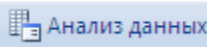
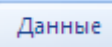
II) Наблюдения -

1) состав кластеров ЮЗ: К5 (С_20-22, Поросино), К1-4 (С_1-19, Зоркальцево), рис. 2.127. В данном случае объекты исследований расположены близко друг от друга в одном секторе (юго-западный), что и сказывается на специфике интерпретации.

2) однородность кластеров ЮЗ по каждому фактору на основании рис. 2.132 и табл.2.10. Юго-западный сектор с близко расположенными объектами и геохимической спецификой дают возможность сравнивать результаты с другими секторами.

3) территориальная неоднородность эколого-геохимических измерений по трем уровням – секторам зон влияния Сибирского химического комбината: юго-западный (ЮЗ) сектор (измерения №1–19 на территории с. Зоркальцево в П₁ и №81–83 - с. Поросино), юго-восточный (ЮВ) сектор (измерения №20–80 на территории г. Томска) и северо-восточный (СВ) сектор (измерения на территории сс. Георгиевка, Наумовка и Надежда №84–91). Территориальная расположенность секторов вызвана поступлением загрязняющих веществ от мощного источника (СХК) согласно розе ветров. Интерпретация результатов учитывается с учетом данных факторов и выделяет северо-восточный сектор, как наиболее загрязненный.

2.9. Средства статистического анализа данных в системе Excel

В системе Excel 2007 представлено большое число статистических функций. Некоторые из них являются встроенными, другие доступны только после загрузки пакета анализа и предназначены для решения сложных статистических задач с помощью подходящей статистической макروفункции. Средства пакета анализа данных доступны через команду  меню  (рис.2.162).

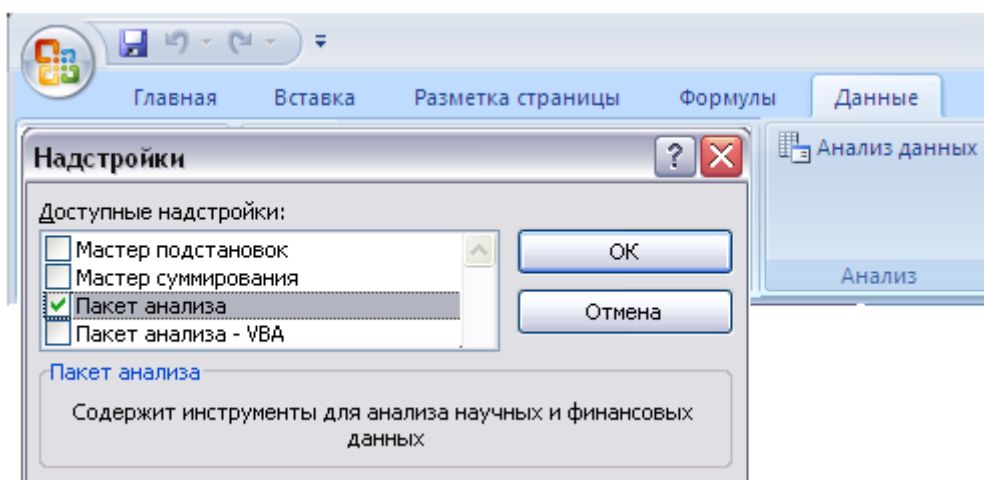

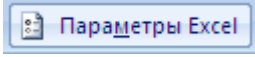
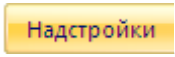
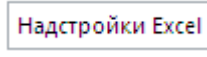
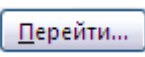

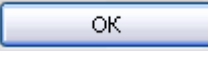


Рис.2.162. Загрузка пакета  в системе Excel

Если этой команды нет, необходимо нажать кнопку , затем щелкнуть , выбрать , далее  и нажать кнопку . В окне  установить флажок напротив «**Пакет анализа**» и нажать  (рис.2.162).

Продemonстрируем возможности системы Excel при поэтапном проведении ССА результатов содержаний микроэлементов, полученных при проведении эколого-геохимических исследованиях разными методами, на уже рассмотренном в пакета **STATISTICA** примере (рис.2.1): ССА содержания микроэлементов в почве территории Томского региона, полученного методами ИНАА и ISP.

На предварительном этапе создается файл (с названием, напр., ИНАА-ISP--.xls) с базой данных (см. Приложения П₁ и рис.2.163).

На **этапе I** для вычисления стандартного набора оценок числовых характеристик содержаний химических элементов в ассоциациях по двум методам в активизированной системе Excel необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль **Данные** (рис.2.163).

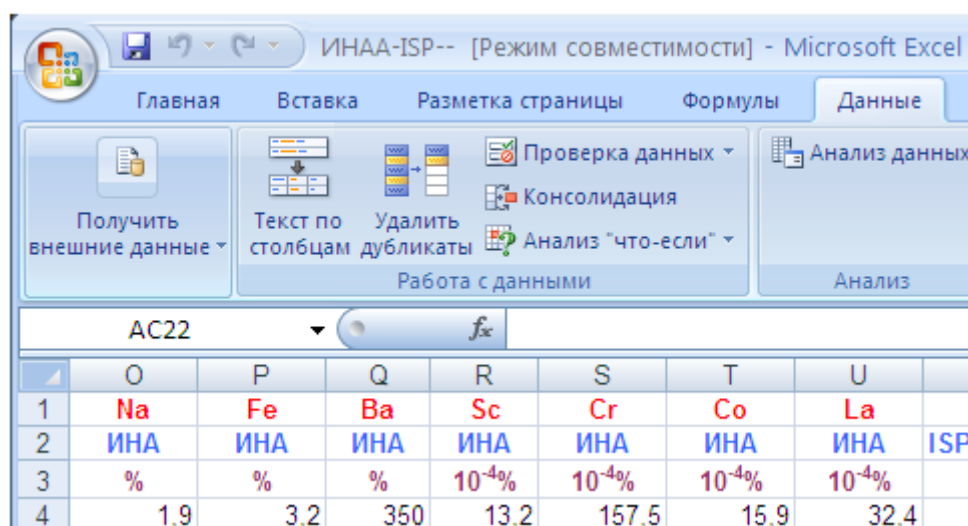


Рис.2.163. Запуск модуля **Данные** в системе Excel

Шаг 2. В стартовой панели модуля нажмите **Анализ данных**

Шаг 3. Далее выберите *Описательная статистика* (рис.2.164)

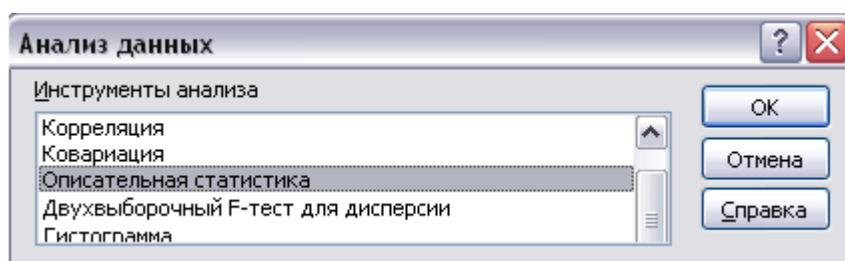


Рис.2.164. Окно **Анализ данных** для выбора инструментов анализа

Шаг 4. В открывшемся окне выберите переменные базы данных, полученных методом ИНАА, указанием номеров левой верхней и правой нижней ячеек соответствующего массива и оформите параметры вывода так, как это показано на рис.2.165.

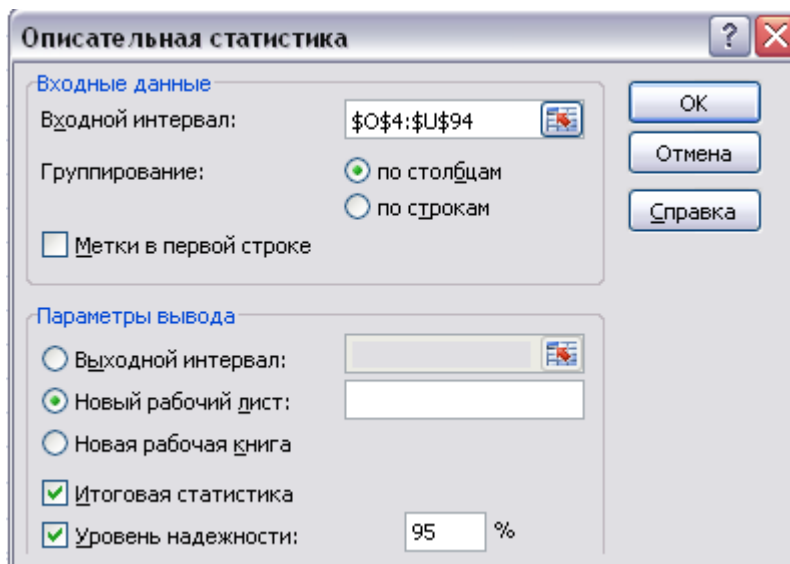


Рис.2.165. Окно **Описательная статистика**

Следующая таблица появится на экране (табл.2.11):

Таблица 2.11

Числовые характеристики содержаний микроэлементов в почве, полученных методом ИНАА

	Na	Fe	Ba	Sc	Cr	Co	La
	ИНА	ИНА	ИНА	ИНА	ИНА	ИНА	ИНА
Среднее m	1,333	3,196	501,1	12,08	127,0	15,79	27,03
Ст. ош. σ_m	0,044	0,071	14,403	0,302	5,362	0,733	0,528
Медиана	1,3	3,3	480	12,8	117,4	15,3	27,5
Мода	1,3	3,2	400	12,8	93,6	14,3	30,8
Ст.откл. S	0,418	0,676	137,394	2,880	51,150	6,990	5,040
Дисперсия	0,175	0,457	18877,1	8,293	2616,31	48,855	25,398
Эксцесс	-0,867	0,424	0,812	-0,281	6,261	27,206	0,082
Асимметрия	0,312	-0,767	0,883	-0,758	2,125	4,231	-0,206
Интервал	1,8	3,3	710	11,4	307,7	60,7	27,1
Минимум	0,5	1,5	260	5,4	46,5	3,7	14
Максимум	2,3	4,8	970	16,8	354,2	64,4	41,1
Сумма	121,3	290,8	45600	1098,9	11557,5	1437,3	2459,3
Счет n	91	91	91	91	91	91	91
Ур.надеж. ε	0,087	0,141	28,614	0,600	10,652	1,456	1,050

Табл.2.11, полученная в системе Excel, аналогична табл.2.1, полученной в системе STATISTICA 6.1. В последней строке табл.2.11 указан 95,0% уровень надежности ε – радиус 95% доверительного интервала (см. рис.1.17 для \mathbf{La}_{II}). Следует заметить, что система Excel оценивает выборки как одномодальные в отличие от системы STATISTICA (выборки и \mathbf{Cr}_{II} в \mathbf{La}_{II} табл.2.1). Еще одно отличие состоит в том, что система Excel не рассчитывает стандартные ошибки σ_A , σ_E , используемые в критерии соответствия эмпирического распределения нормальному теоретическому (1.22).



Аналогичным образом вычисляются оценки числовых характеристик содержаний микроэлементов в почве, полученных методом ISP (табл.2.12).

Таблица 2.12

Числовые характеристики содержаний микроэлементов в почве, полученных методом ISP

	Na	Fe	Ba	Sc	Cr	Co	La
	ISP	ISP	ISP	ISP	ISP	ISP	ISP
Среднее	1,786	3,041	406,63	9,180	117,51	18,647	22,207
Ст. ошибка	0,051	0,070	9,682	0,255	5,189	0,926	0,725
Медиана	1,6973	3,0476	387	9	104	19	24
Мода	1,99	2,99	380	10	80	18	20
Ст. откл.	0,486	0,668	92,365	2,433	49,503	8,835	6,920
Дисперсия	0,237	0,446	8531,2	5,921	2450,56	78,058	47,887
Экссесс	-0,186	1,731	9,553	0,077	10,503	13,037	0,273
Асимметрия	0,645	-0,570	2,484	0,097	2,761	2,674	-0,709
Интервал	2,083	3,58	638	11,3	297,1	62	31,1
Минимум	0,877	0,99	245	3	62,9	3	2
Максимум	2,96	4,57	883	14,3	360	65	33,1
Сумма	162,51	276,77	37003	835,4	10693,6	1696,9	2020,8
Счет	91	91	91	91	91	91	91
Ур.надеж. ε	0,101	0,139	19,236	0,507	10,310	1,840	1,441

Если вычисленные числовые характеристики носят промежуточный характер (используются в дальнейших расчетах), то отдельные их значения можно вычислить в указанных ячейках с помощью встроенных статистических функций. Например, для вычисления среднего содержания микроэлемента \mathbf{Na}_{II} можно выполнить последовательность следующих действий:

Шаг 1. Выберите место для результата (ячейка O95) и в головном меню с помощью кнопки  запустите  (рис.2.166).

Шаг 2. В открывшемся окне выделите категорию **Статистические**, выберите функцию **СРЗНАЧ** (рис.2.166) и нажмите **ОК**.

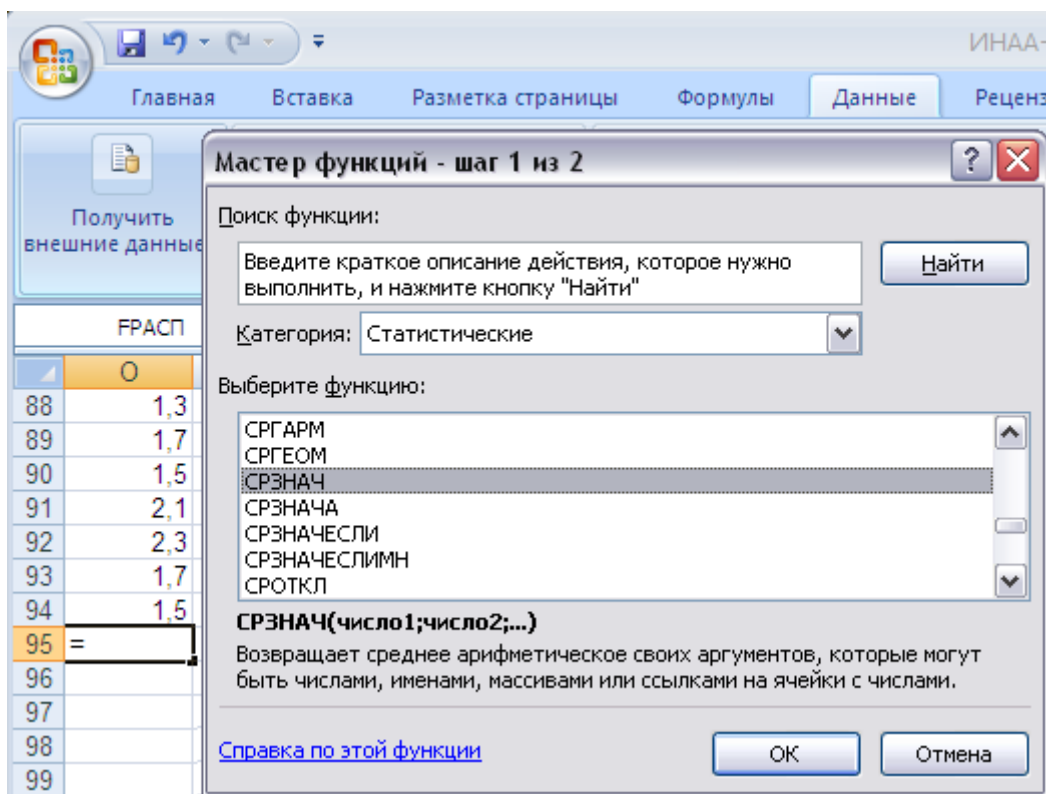


Рис.2.166. Открытие окна **Мастер функций**

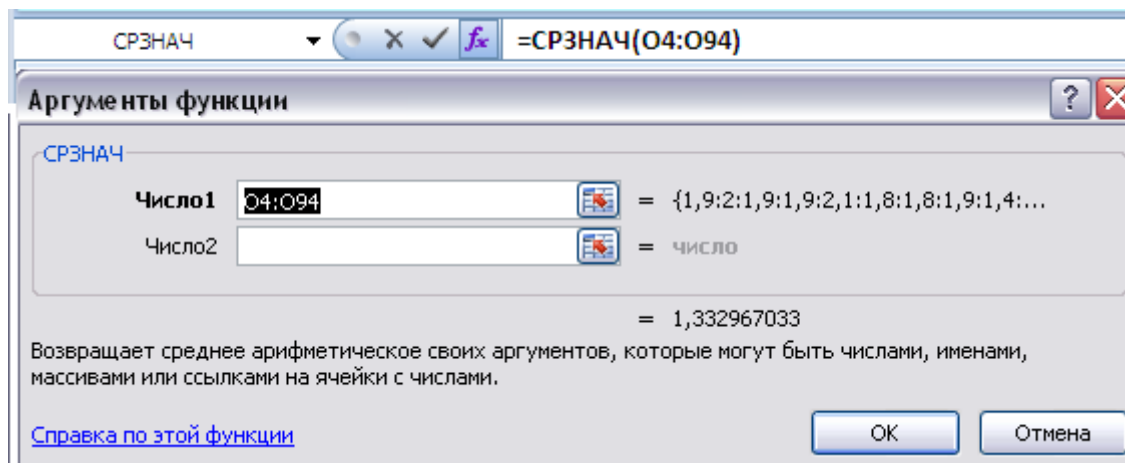



Рис.2.167. Окно **Аргументы функции** **СРЗНАЧ**

Шаг 3. В окне **Аргументы функции** (см. рис.2.167) система выдаст вычисленное значение среднего, а после нажатия **ОК** поместит его в указанную ячейку O95.

Рассмотрим далее построение гистограммы содержания микро-элемента Na_{II} , аналогичной рис.2.12. В пошаговом исполнении эта процедура может быть оформлена следующим образом:

Шаг 1. Запустите **Мастер функций** в головном меню с помощью кнопки  (рис.2.166).

Шаг 2. В открывшемся окне выделите категорию **Статистические**, выберите функцию *ЧАСТОТА* и нажмите ОК.

Шаг 3. Далее выберите **Аргументы функции** (см. рис.2.168), задавая границы ячеек O4:O94 массива данных Na_{II} и границы ячеек AD4:AD11 массива интервалов (8 внутренних границ, разбивающих интервал значений 0,5 – 2,3 для Na_{II} (табл.2.11) равномерно на 9 интервалов), после чего система выдаст здесь же вычисленный частотный ряд (табл.1.4) для Na_{II} и поместит его ниже массива интервалов в ячейки AD12:AD20.

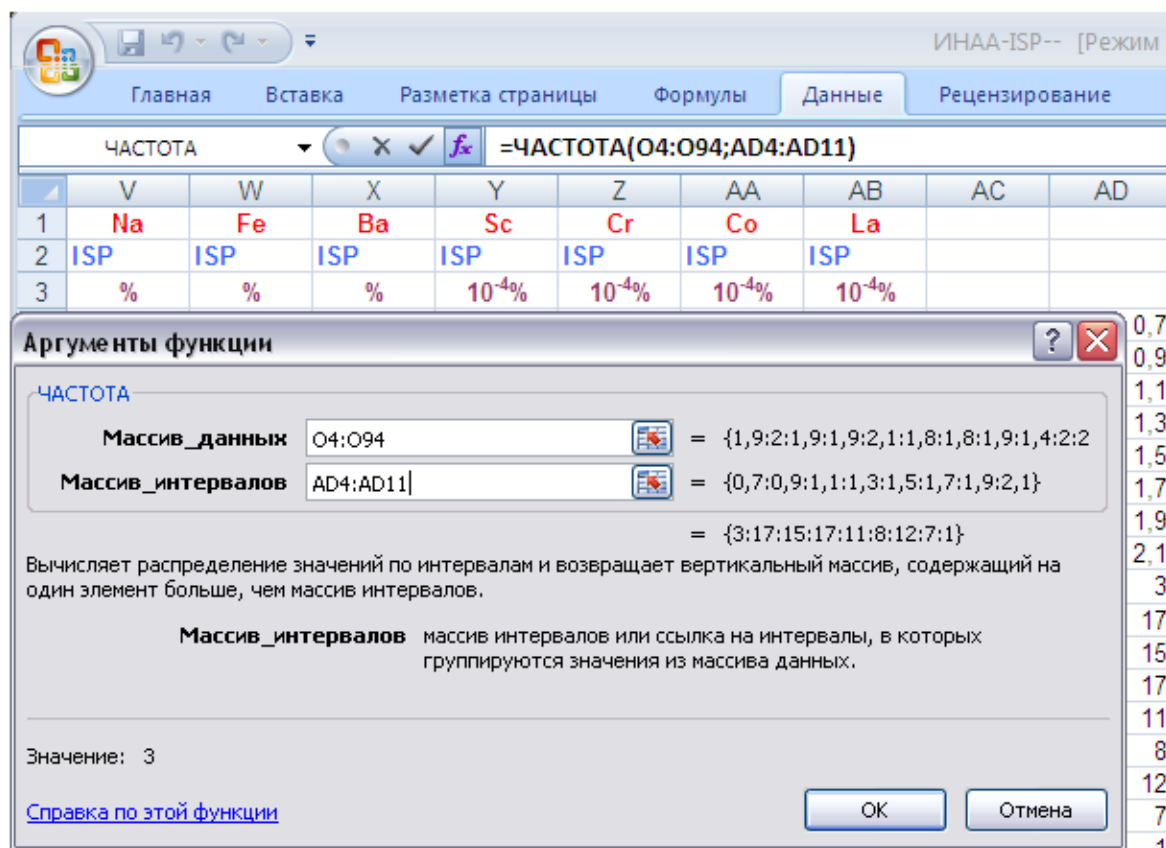



Рис.2.168. Окно **Аргументы функции** *ЧАСТОТА*

Шаг 4. Выделите массив ячеек AD12:AD20. Далее запустите в головном меню в модуле **Вставка**  и

выберите цилиндрический тип  (рис.2.169) и щелкните по ней. После чего в режиме **Работа с диаграммами** появится гистограмма (рис.2.170)

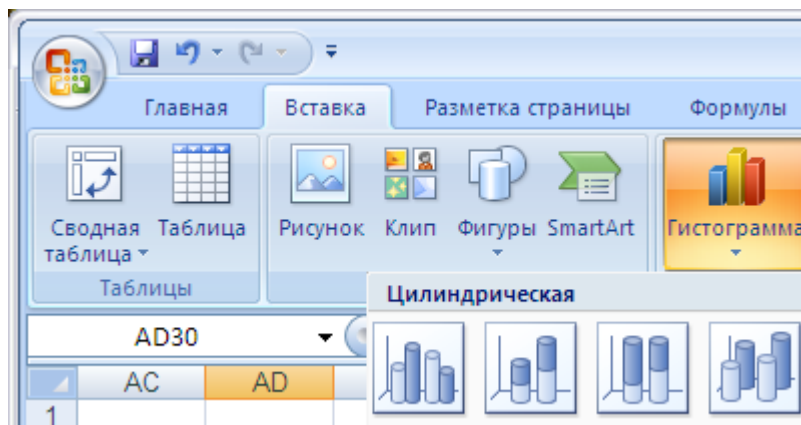


Рис.2.169. Окно выбора типа гистограммы

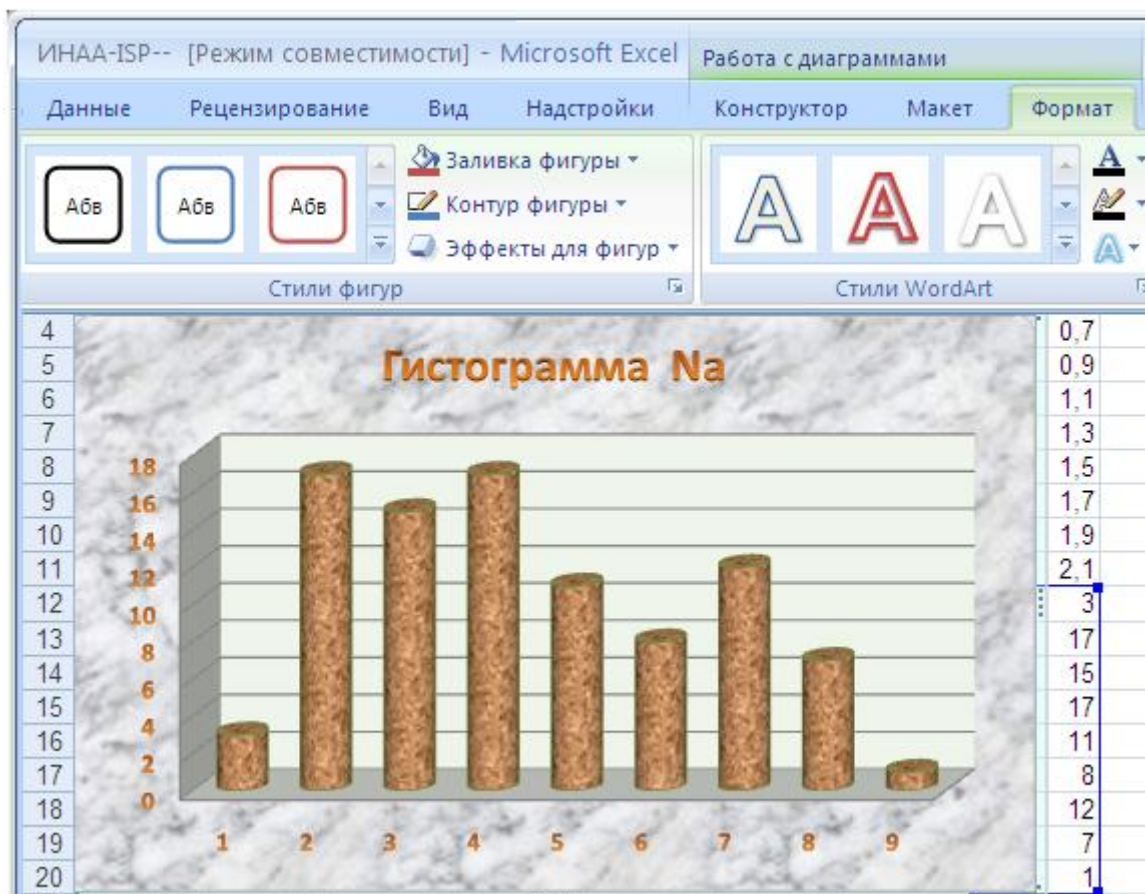


Рис.2.170. Гистограмма №_и (вариант №1)

Шаг 5. В режиме **Работа с диаграммами** имеются различные средства по корректировке формы гистограммы (рис.2.170' и 2.170'').

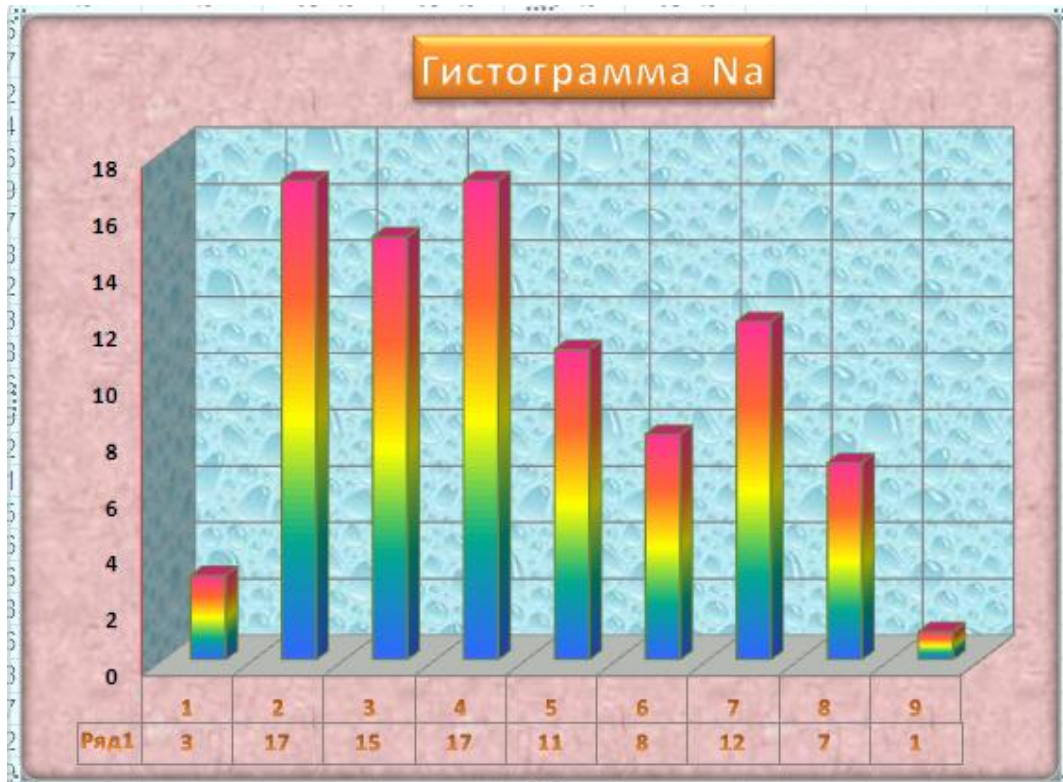


Рис.2.170'. Гистограмма Na_{II} (вариант №2)

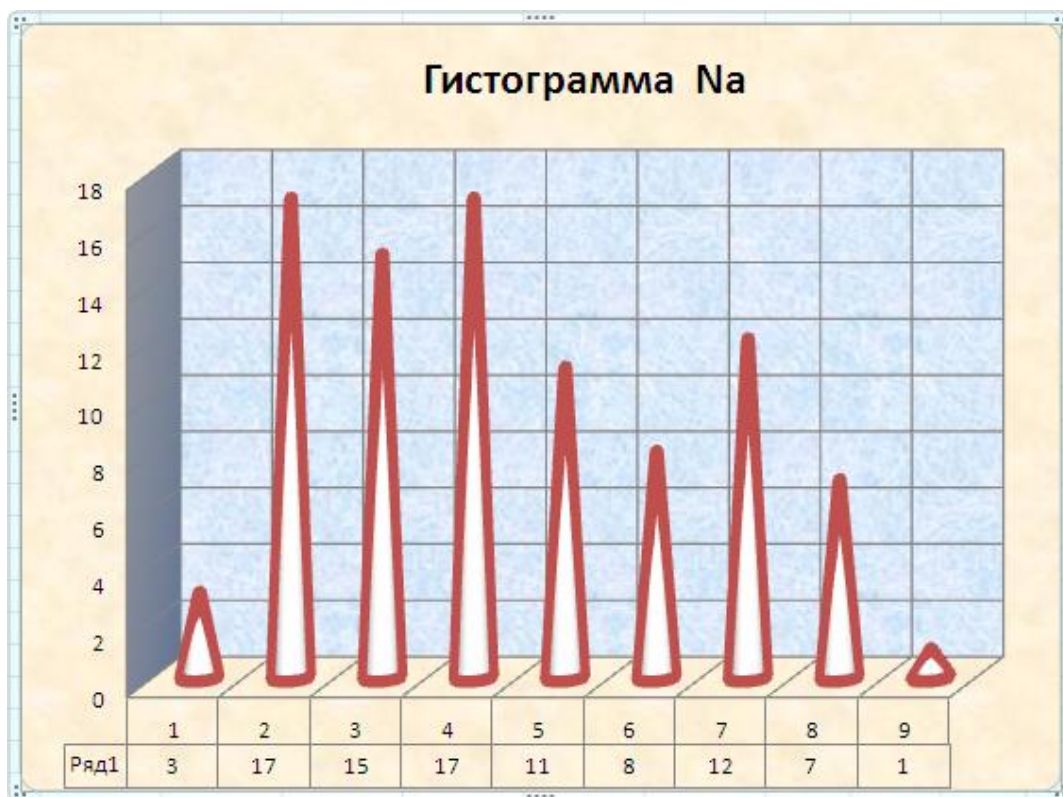


Рис.2.170''. Гистограмма Na_{II} (вариант №3)

Более компактный вариант построения гистограммы реализуется с помощью макروفункции *Гистограмма*: (рис. 2.163-164), после выбора которой в открывшемся окне выберите переменную **Na_{II}** базы данных, указав номера левой верхней и правой нижней ячеек соответствующего массива, и оформите параметры вывода, как это показано на рис.2.171. В результате на экране появится гистограмма содержания микроэлемента **Na_{II}**, которую можно корректировать в режиме Работа с диаграммами (рис.2. 171).



Рис.2.171. Окно Гистограмма и гистограмма содержания микроэлемента **Na_{II}** с автоматическим разбиением интервала значений 0,5 – 2,3

При этом система выполнила автоматическое разбиение интервала значений 0,5 – 2,3 для **Na_{II}** равномерно на 10 интервалов. Заметим, что система допускает также задание пользователем массива интервалов (аналогично рис. 2.168) в окне интервал карманов.

В первой части **этапа II** для корректного выбора критерия сравнения двух выборок необходимо предварительно проверить гипотезу о нормальном распределении содержаний химических элементов. К сожалению в системе Excel нет подходящей статистической макروفунк-

ции по аналогии с пакетом STATISTICA. Поэтому в данном случае приходится поэтапно использовать ряд статистических функций, последней из которых является *ХИ2ТЕСТ*(*фактический интервал; ожидаемый интервал*), вычисляющая уровень значимости отличия фактического (экспериментального) распределения сразу нескольких переменных от соответствующих ожидаемых (теоретических).

Статистическое значение критерия вычисляется по формуле

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{[A_{ij} - E_{ij}]^2}{E_{ij}},$$

где k - число строк (число интервалов),

m - число столбцов (число переменных),

A_{ij} - фактическая частота в i -ой строке, j -ом столбце,

E_{ij} - ожидаемая частота в i -ой строке, j -ом столбце.

Построенная случайная величина распределена по закону хи-квадрат с $df = (k-1)(m-1)$ числом степеней свободы. Таким образом, этот совместный критерий ($m > 1$) обладает тем недостатком, что может оценивать незначимым отличие распределения совместной выборки (A_{ij}) от нормального закона в том случае, когда распределение одной переменной отличается от нормального закона незначимо, другой – значимо. Для результативного исследования достаточно, чтобы значимо отличалось от нормального закона распределение хотя бы одной переменной по критерию Пирсона (1.19). Еще одним недостатком совместного критерия по наличию в df множителя $(k-1)$ является то, что он не учитывает число неизвестных параметров ожидаемого нормального закона ($L=2$ по критерию (1.19)), что изменяет число степеней свободы и может привести к некорректной оценке вычисляемого уровня значимости отличия фактического распределения от ожидаемого по нормальному закону. Наконец, совместный критерий в системе Excel в отличие от пакета STATISTICA совершенно равнодушен к требованию $E_{ij} \geq 5$, нарушение которого приводит посредством объединения соответствующих интервалов к изменению как самого критерия, так и числа степеней свободы и, как следствие, к изменению вычисляемого уровня значимости отличия экспериментального распределения от теоретического нормального закона.

Рассмотрим проверку гипотезы о нормальном распределении содержания химического элемента La по методам ИНАА и ISP с поэтапным использованием ряда статистических функций и промежуточных

вычислений в системе Excel. Рабочее поле вычислений по формуле (1.19) представлено на рис.2.172 и с помощью *XI2TEST* на рис.2.173.

	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
1	La	27,03					La	22,21				
2	ИНА	5,04					ISP	6,92				
3	10⁻⁴%						10⁻⁴%					
4	32,4	5					28	5				
5	34,2	10					26,2	10				
6	32,4	15					24,9	15				
7	35,3	20					33,1	20				
8	34,9	25					27,5	25				
9	30,8	30					26,2	30				
10	30,1	35					27	35				
11	30,8	40		0			27,8	40		0		
12	29,6	0	6E-04	6E-06			26,4	3	0,587	0,006		
13	30,5	0	0,033	4E-04			27,6	2	2,95	0,039		
14	30,2	1	0,742	0,009	1	0,775	28,1	9	10,01	0,149	14	13,54
15	28,6	8	6,656	0,082	8	6,656	27,3	23	20,57	0,375	23	20,57
16	30,6	19	23,86	0,344	19	23,86	26,6	19	25,65	0,657	19	25,65
17	30,1	35	34,45	0,722	35	34,45	31,1	25	19,4	0,87	25	19,4
18	34,5	25	20,09	0,943	28	25,25	30,8	10	8,899	0,968	10	11,83
19	30,8	2	4,71	0,995			29,7	0	2,473	0,995		
20	30,7	1	0,457	1			28,9	0	0,461	1		
21	29,1						31,4					
22	27,9	6E-04					30,2	9,919				
23	28,1	0,033					16	0,306				
24	16,2	0,09			0,066		16	0,101			0,015	
25	18,9	0,271			0,271		9	0,286			0,286	
26	28,6	0,991			0,991		13	1,724			1,724	
27	22,4	0,009			0,009		14	1,616			1,616	
28	20,6	1,202			0,299		13	0,136			0,284	
29	24,4	1,559					8	2,473				
30	20,5	0,646					13	0,461				
31	21,8						16					
32	19,8	4,802			1,635		15	17,02			3,926	
33	26	0,569			0,441		17	0,009			0,14	
34	23						20					

Рис.2.172. Рабочее поле проверки гипотезы о нормальном распределении содержаний химического элемента La по методам ИНАА и ISP по формуле (1.19)

На рис.2.172 фактические измерения содержаний химического элемента La по методу ИНАА оформлены в ячейках AI4:AI94 и по методу ISP в ячейках AO4:AO94.

Для проверки гипотезы о нормальном распределении содержания химического элемента по формуле (1.19) на примере $La_{и}$ (La_{ISP} аналогично) можно выполнить последовательность следующих действий:

Шаг 1. Оформим массив интервалов в ячейках AJ4:AJ11 и вычислим фактические частоты (рис.2.168) с помощью функции

$$ЧАСТОТА(AI4:AI94;AJ4:AJ11),$$

которые разместятся в ячейках AJ12:AJ20.

Шаг 2. Для вычисления ожидаемых (по нормальному закону) частот предварительно вычислим в ячейке AJ1 среднее (рис.2.166-167) и в ячейке AJ2 стандартное отклонение с помощью функций соответственно

$$\begin{aligned} &CP3HAЧ(AI4:AI94) \\ &СТАНДОТКЛОН(AI4:AI94) \end{aligned}$$

Шаг 3. Далее вычислим в ячейках AL12:AL19 значения функции распределения $F(x)$ по нормальному закону по массиву интервалов (перебирая ячейки AJ4:AJ11) с помощью функции

$$\begin{aligned} &НОРМРАСП(AJ4;AJ1;AJ2;ИСТИНА) \text{ или} \\ &НОРМРАСП(AJ4;27,02527;5,039656;ИСТИНА). \end{aligned}$$

Крайние значения (0 в ячейку AL11 и 1 в ячейку AL20) добавляем согласно свойствам $F(x)$ (1.4).

Шаг 4. Затем вычислим в ячейках АК12:АК20 значения теоретических частот np_i для (1.19) с помощью $F(x)$ по формуле, например в ячейку АК12,

$$(AL12-AL11)*91.$$

Шаг 5. После этого вычислим в ячейках AJ22:AJ30 значения χ_i^2 в формуле (1.19), например в ячейку АК12

$$(AJ12-АК12)^2/АК12.$$

Перебрав таким образом все ячейки AJ22:AJ30, вычислим в ячейке AJ32 значение критерия Пирсона (1.19) по формуле

$$СУММ(AJ22:AJ30).$$

Шаг 6. Соответствующее вычисленному критерию Пирсона значение уровня значимости вычисляем в ячейке AJ33 с помощью функции

$$ХИ2РАСП(AJ32;6) = 0,569.$$

Аналогично в столбцах AM-AN оформлен вариант для $La_{и}$ с учетом требования $E_{ij} \geq 5$, когда три крайние интервала объединены в один.

Таким образом скорректированный уровень значимости равен 0,441.

Следовательно можно считать $La_{и}$ нормально распределенным.

	AU	AV	AW	AX	AY	AZ
1		La	La	La	La	
2		ИНА	ISP	ИНА	ISP	
3		0	3			
4		0	2			
5		1	9	1	14	
6		8	23	8	23	
7		19	19	19	19	
8		35	25	35	25	
9		25	10	28	10	
10		2	0			
11		1	0			
12		0,0006	0,587			
13		0,0326	2,9502			
14		0,7415	10,008	0,7747	13,545	
15		6,6563	20,572	6,6563	20,572	
16	0,005	23,863	25,649	23,863	25,649	0,234
17		34,453	19,4	34,453	19,4	
18		20,086	8,8991	25,253	11,834	
19		4,7102	2,4735			
20		0,4567	0,461			
21						
22		0,0006	9,919			
23		0,0326	0,306			
24		0,0901	0,1014	0,0655	0,0153	
25	21,83	0,2713	0,2865	0,2713	0,2865	5,561
26	0,005	0,9911	1,7237	0,9911	1,7237	0,234
27	0,001	0,0087	1,6163	0,0087	1,6163	0,062
28		1,2021	0,1362	0,2988	0,2841	
29		1,5595	2,4735			
30		0,6462	0,461			
31						
32		4,8021	17,02	1,635	3,9259	
33		0,5694	0,009	0,441	0,1404	
34						

Рис.2.173. Рабочее поле проверки гипотезы о нормальном распределении содержаний химического элемента La по методам ИНАА и ISP с помощью *ХИ2ТЕСТ*

Шаг 7. Оформим совместное содержания химического элемента La по методам ИНАА и ISP (рис.2. 173) вычисленными ранее фактическими и ожидаемыми частотами: AJ12:AJ20 в ячейках AV3:AV11, AP12:AP20 в ячейках AW3:AW11, AM14:AM18 в ячейках AX5:AX9, AS14:AS18 в ячейках AY5:AY9, AK12:AK20 в ячейках AV12:AV20, AQ12:AQ20 в ячейках AW12:AW20, AN14:AN18 в ячейках AX14:AX18, AT14:AT18 в ячейках AY14:AY18.

Шаг 8. Для проверки гипотезы о нормальном распределении содержания химического элемента с помощью *XI2TECT* на совместном примере La_{II} и La_{ISP} в ячейке AU16 вычислим уровень значимости отличия фактического распределения La_{II} и La_{ISP} от нормального распределения по 9-ти интервалам ($df = 8$)

$$XI2TECT(AV3:AW11;AV12:AW20),$$

а в ячейке AZ16 - по 5-ти интервалам ($df = 4$)

$$XI2TECT(AX5:AY9;AX14:AY18).$$

Таким образом, за счет учета требования $E_{ij} \geq 5$, уровень значимости изменился от значимого 0,005 до незначимого 0,234.

Шаг 9. Проконтролируем *XI2TECT*, вычисляя объединенный критерий непосредственно (аналогично (1.19) в рис.2.172). Например, в ячейке AW22 по формуле

$$(AW3-AW12)^2/AW12 \text{ и т.д.,}$$

в ячейке AW32 по формуле

$$СУММ(AW22:AW30) \text{ и т.д.,}$$

в ячейке AU25 вычисляем объединенный критерий по 9-ти интервалам $AV32+AW32$,

а в ячейке AZ25 вычисляем объединенный критерий по 5-ти интервалам $AX32+AY32$.

Наконец находим соответствующие уровни значимости:

в ячейке AU26 по 9-ти интервалам ($df = 8$) с помощью функции

$$XI2РАСП(AU25;8) = 0,005,$$

а в ячейке AZ26 по 5-ти интервалам ($df = 4$) с помощью функции

$$XI2РАСП(AZ25;4) = 0,234.$$

Шаг 10. Скорректируем вычисленные уровни значимости объединенного критерия за счет учета числа неизвестных параметров ожидаемого нормального закона, что приводит к уменьшению df на 2:

в ячейке AU27 по 9-ти интервалам ($df = 6$) с помощью функции

$$XI2РАСП(AU25;6) = 0,001,$$

а в ячейке AZ27 по 5-ти интервалам ($df = 2$) с помощью функции

$$XI2РАСП(AZ25;2) = 0,062.$$

Таким образом, скорректированный уровень значимости объединенного критерия равен 0,062. Иными словами, распределения содержаний химического элемента La по совокупности методов ИНАА и ISP слабо значимо отличаются от нормального закона.

Наглядное представление различий экспериментального распределения содержаний химического элемента La по методам ИНАА и ISP от нормального закона дают гладкие графики фактических (ИНАА-Э, ISP-Э) и ожидаемых (ИНАА-Т, ISP-Т) частот (рис.2.172-173), изображенные на рис. 2. 174.



Рис.2.174. Составной гладкий полигон частот содержания микроэлемента La по методам ИНАА и ISP

Во второй части **этапа II** для сравнения содержаний химических элементов, полученных при проведении эколого-геохимических исследований разными методами, можно применить в Excel только критерии Фишера (1.23) и Стьюдента (1.24-25) в случае нормально распределенных содержаний химических элементов.

Для начала сравнение выборок можно провести графически: аналогично рис.2.81 по составным гистограммам (рис. 2.175).



Рис.2.175. Составная гистограмма содержания микроэлемента La по методам ИНАА и ISP

Для проведения аналитической оценки различия содержаний химических элементов $La_{И}$ и La_{ISP} по критерию Стьюдента необходимо выполнить в Excel последовательность следующих действий:

Шаг 1. Запустите в главном меню модуль (рис.2.163).

Шаг 2. В стартовой панели модуля нажмите .

Шаг 3. Далее выберите *Двухвыборочный F-тест для дисперсий* (рис.2.176)

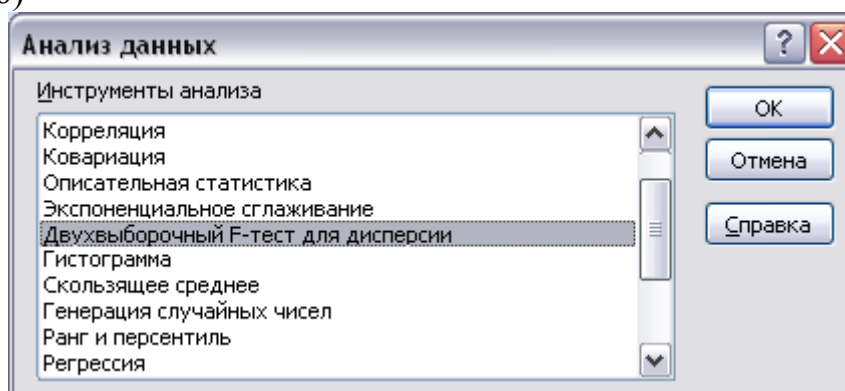


Рис.2. 176. Окно для выбора *F-тест*

Шаг 4. В открывшемся окне выберите переменные $La_{И}$ и La_{ISP} , как это показано на рис.2.177.

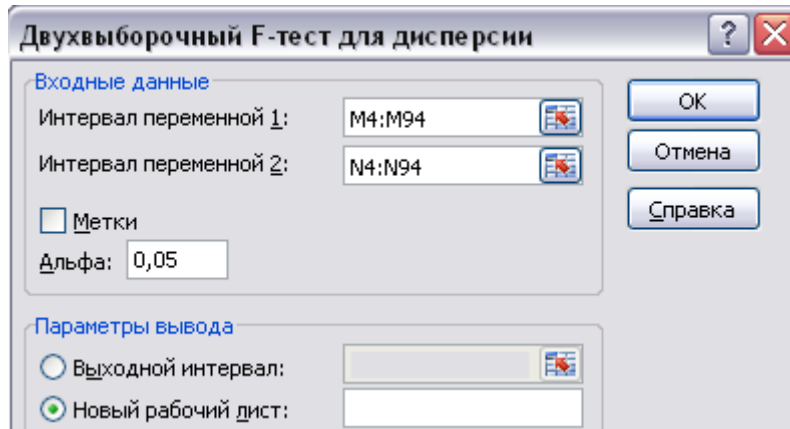


Рис.2.177. Окно **Двухвыборочный F-тест для дисперсии**

Следующая таблица появится на экране (табл.2.13):

Таблица 2.13

Двухвыборочный F-тест для дисперсий переменных La_{II} и La_{ISP}

	A	B	C
4	Среднее	27,02527473	22,20659341
5	Дисперсия	25,39813187	47,88706716
6	Наблюдения	91	91
7	df	90	90
8	F	0,530375598	
9	P(F<=f) одностороннее	0,001463628	
10	F критическое одностороннее	0,705672307	

Как следует из табл. 2.12 ($p_F = 0,001463628$), дисперсии переменных La_{II} и La_{ISP} различаются значимо.

Шаг 5. Еще раз загружаем 

Шаг 6. Далее выберите *Двухвыборочный t-тест с различными дисперсиями* (рис.2.178)

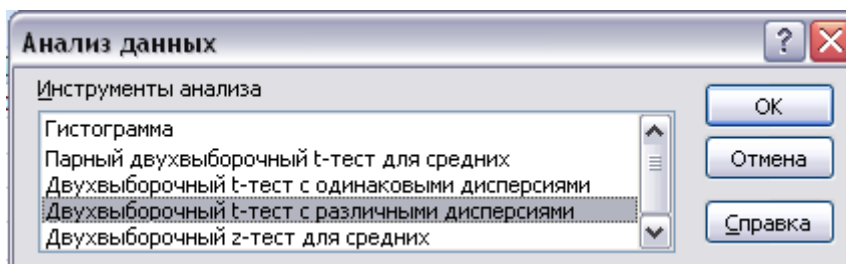
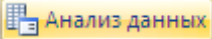


Рис.2.178. Окно  для выбора *t-тест*

Шаг 7. В открывшемся окне выберите переменные La_{II} и La_{ISP} , как это показано на рис.2.177. На экране появится таблица (табл.2.14):

Таблица 2.14

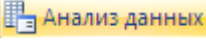
Двухвыборочный t-тест для средних переменных $L_{aИ}$ и L_{aISP}

	A	B	C
1	Двухвыборочный t-тест с различными дисперсиями		
2			
3		Переменная 1	Переменная 2
4	Среднее	27,02527473	22,20659341
5	Дисперсия	25,39813187	47,88706716
6	Наблюдения	91	91
7	Гипотетическая разность средних	0	
8	df	165	
9	t-статистика	5,369585546	
10	P(T<=t) одностороннее	1,32199E-07	
11	t критическое одностороннее	1,654140977	
12	P(T<=t) двухстороннее	2,64398E-07	
13	t критическое двухстороннее	1,974445593	

Согласно табл.2.14 содержания микроэлемента L_a , полученные методами ИНАА и ISP, различаются высоко значимо ($p < 0,001$).

Следует заметить, что Excel реализованы, согласно рис. 2.156, все варианты критериев сравнения средних нормально распределенных выборок: двухвыборочный z-тест для средних (1.24), двухвыборочный t-тест с одинаковыми дисперсиями (1.25) и двухвыборочный t-тест с различными дисперсиями (1.25)*. Если распределения выборок далеки от нормального закона, то вопрос о различии выборок в системе Excel остается открытым в виду отсутствия реализации соответствующих непараметрических критериев.

В первой части **этапа III** вычислим корреляционную матрицу одной из двух ассоциаций, например, по методу ИНАА:

Шаг 1. Запустите , затем *Корреляция* (рис.2.176)

Шаг 2. В открывшемся окне выберите переменные (рис.2.179).

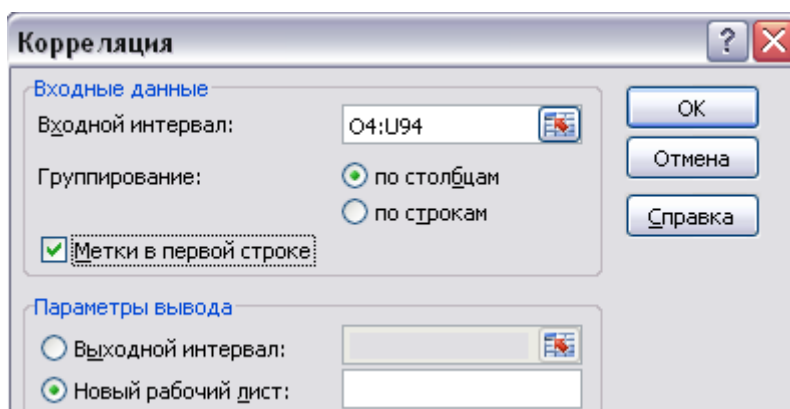


Рис.2.179. Окно **Корреляция**

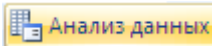
Следующая таблица появится на экране (табл.2.15), аналогичная табл.2.5 :

Таблица 2.15

Корреляционная матрица выборки содержаний микроэлементов в почво-грунтах, полученных методом ИНАА.

	Na _и	Fe _и	Ba _и	Sc _и	Cr _и	Co _и	La _и
Na _и	1,000						
Fe _и	0,142	1,000					
Ba _и	-0,336	0,110	1,000				
Sc _и	0,372	0,871	-0,085	1,000			
Cr _и	0,159	0,241	0,140	0,182	1,000		
Co _и	0,028	0,379	0,332	0,240	0,594	1,000	
La _и	0,439	0,642	-0,187	0,762	0,119	0,142	1,000

Во второй части **этапа III** для построения диаграммы рассеяния и уравнения регрессии для наиболее значимой корреляционной связи Fe и Sc в ассоциации метода ИНАА необходимо выполнить последовательность следующих действий:

Шаг 1. Запустите , затем *Регрессия* (рис.2.176)

Шаг 2. В открывшемся окне выберите переменные, например, Fe (P4:P94) и Sc (R4:R94) по методу ИНАА, как это показано на рис.2.180.

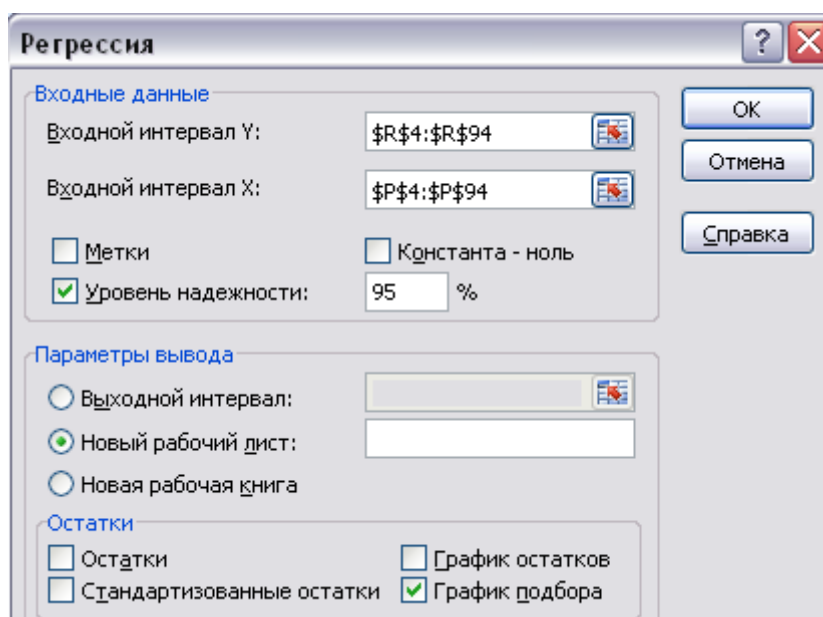


Рис.2.180. Окно 

Следующая сборная таблица появится на экране (табл.2.16):

Таблица 2.16

Регрессионная статистика связи Fe и Sc по методу ИНАА

Регрессионная статистика			Коэфф	Станд. ош.
Множественный R	0,870151			
R-квадрат	0,757163	Y-пересечение	0,234631	0,7264
Нормир. R-квадрат	0,754434	Переменная X		
		1	3,705463	0,2224
Дисперсионный анализ				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	1	565,1213	565,1213	277,501
Остаток	89	181,2455	2,036466	
Итого	90	746,3668		

Также появится на экране следующая диаграмма рассеяния (рис.2.181), аналогичная рис.2.81.

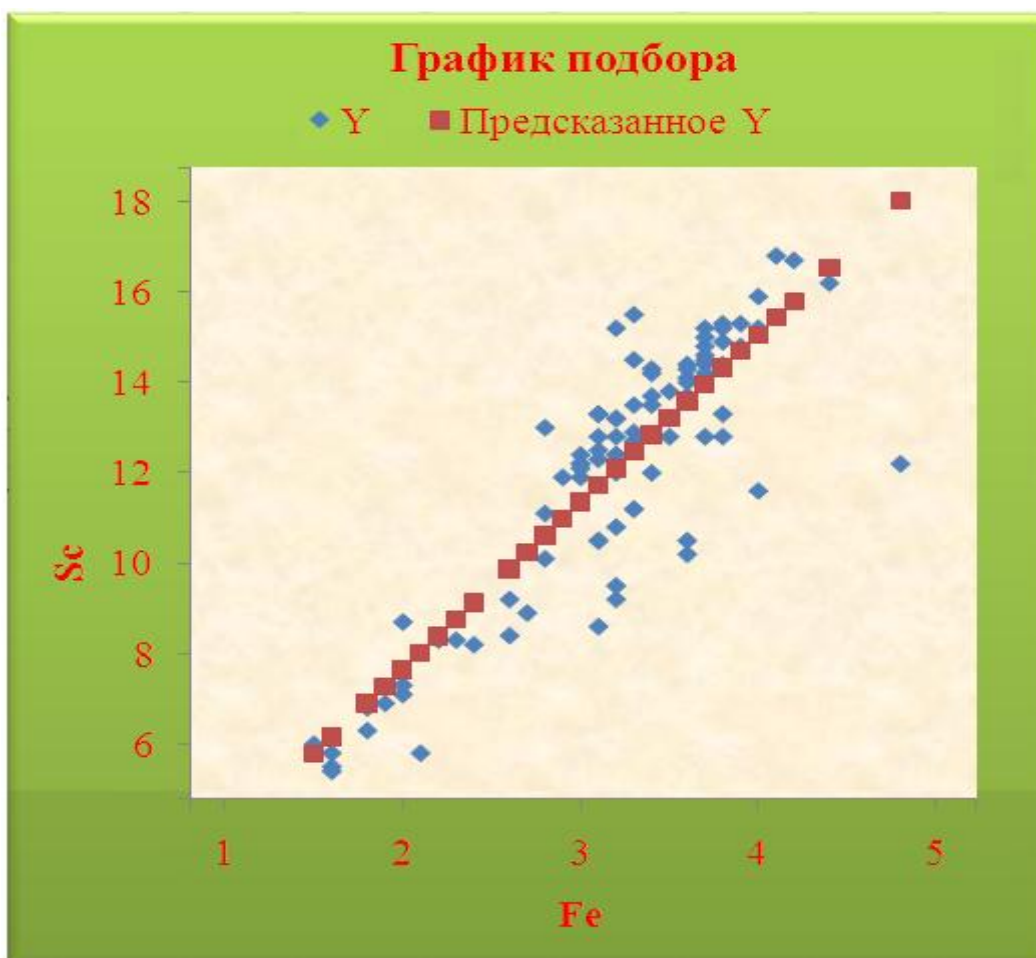


Рис. 2.181. Диаграмма рассеяния зависимости Fe и Sc по методу ИНАА

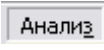
2.10. Модульный анализ данных в системе STATISTICA

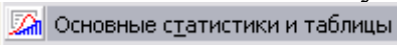
В версии STATISTICA 6.1 реализованы огромные возможности языка **STATISTICA VISUAL BASIC (SVB)**, позволяющие сложные процедуры анализа и графический вывод результатов записывать как макрос или сценарий анализа для дальнейшего многократного использования и редактирования. Макросы представляют собой самостоятельные блоки, которые легко встраиваются в другие приложения.

Различают простые макросы анализа, используемые в одном модуле (например, *Описательные статистики*), и мастер – макросы, позволяющие записывать сценарии сложных исследований, включающих несколько видов анализа с использованием различных модулей системы. Простой макрос записывают сразу после окончания непрерывной процедуры анализа. Записью мастер – макроса можно управлять, то есть в любой момент можно начать запись или приостановить ее. Данная возможность придает системе большую гибкость и позволяет связывать различные виды анализа. Мастер – макрос может быть в дальнейшем отредактирован, сохранен или запущен на исполнение, с целью точного повторения записанной последовательности действий. Таким образом, можно создавать макрос, чтобы затем включить его в окончательный отчет или запускать его для подобного анализа других данных.


Рассмотрим создание макросов на примере ССА эколого-геохимических исследований, приведенном в п.2.1-4.

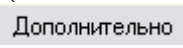
На **этапе I** можно создать макрос описательной статистики (МОС) для вычисления оценок числовых характеристик содержаний химических элементов (п.2.1.1), построения диаграмм размаха (п.2.1.2) и построения гистограмм (см. п. 2.1.3). Для реализации этого сценария анализа в активизированном пакете STATISTICA следует выполнить последовательность следующих действий:



Шаг 1. Запустите в головном меню модуль  (рис.2.2).




Шаг 2. Нажмите  в стартовой панели модуля (рис.2.3).

Шаг 3. В появившемся окне выберите .


Шаг 4. В открывшемся окне нажмите кнопку , выберите переменные базы данных (столбцы 1-14 на рис.2.1) и нажмите ОК.

Шаг 5. Нажмите кнопку  и закажите (поставьте все нужные числовые характеристики (рис.2.4)).

Шаг 6. Нажмите  в окне . На экране появится таблица (табл.2.1).

Шаг 7. Вернитесь в окно  **Описательные статистики** (рис.2.4), нажмите кнопку  **Переменные:**, выберите категориальные переменные (столбцы 16-22 на рис.2.25) и нажмите .

Шаг 8. Нажмите кнопку **Опции** (рис.2.5) и закажите тип графика выбором **Опции для диаграмм размаха**, например, **Среднее/Ст.ош./Ст.отк** (Среднее/стандартная ошибка/ стандартное отклонение).

Шаг 9. Закажите  **Категоризованные диаграммы размаха** в режиме **Катег. графики** и (рис.2.182)

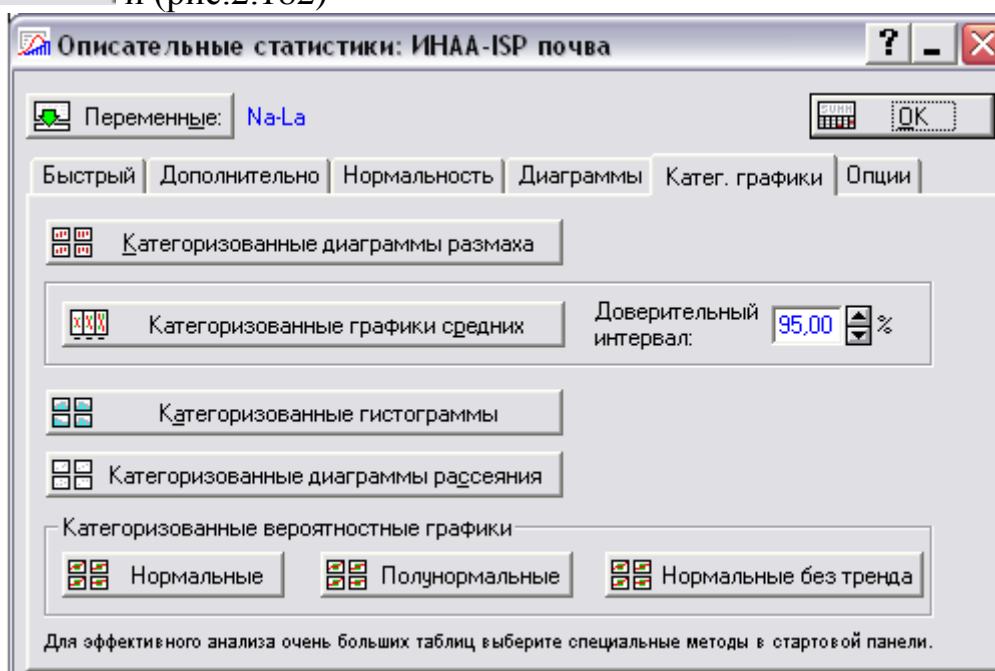


Рис. 2.182. Режим **Катег. графики** в окне  **Описательные статистики:**

Шаг 10. В открывшемся окне (рис.2.183), выберите группирующую переменную (столбец 15) и нажмите ОК.

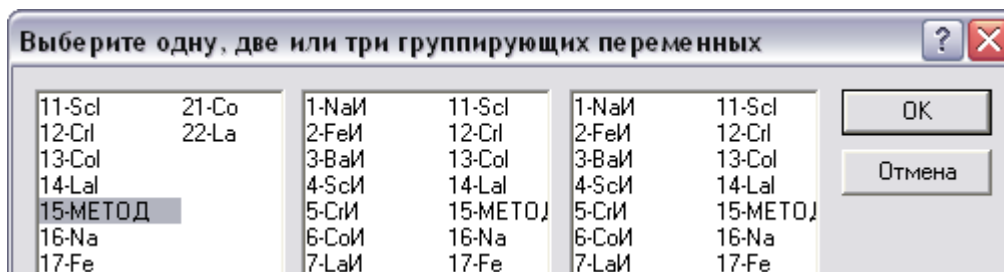



Рис. 2.183. Окно  **Переменные:** для выбора группирующей переменной (столбец 15)

Шаг 11. В открывшемся окне (рис.2.184), выберите коды группирующей переменной (кнопка **Все**) и нажмите **ОК**. На экране появятся составные по методам диаграммы размаха типа Среднее/... для всех микроэлементов (аналогично рис.2.9).

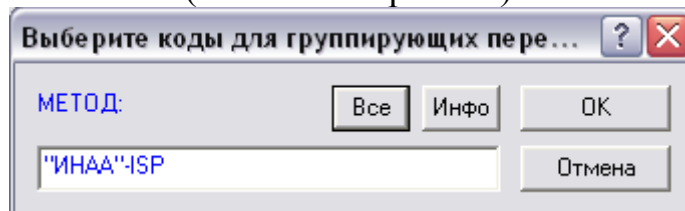


Рис. 2.184. Окно для выбора кодов группирующей переменной (кнопка

Все)

Шаг 12. Вернитесь в окно **Описательные статистики:** (рис.2.182), нажмите кнопку **Катег. графики** и закажите **Категоризованные гистограммы**. Далее повторите последовательно шаги 10-11. На экране появятся составные по методам гистограммы для всех микроэлементов (рис.2.185).

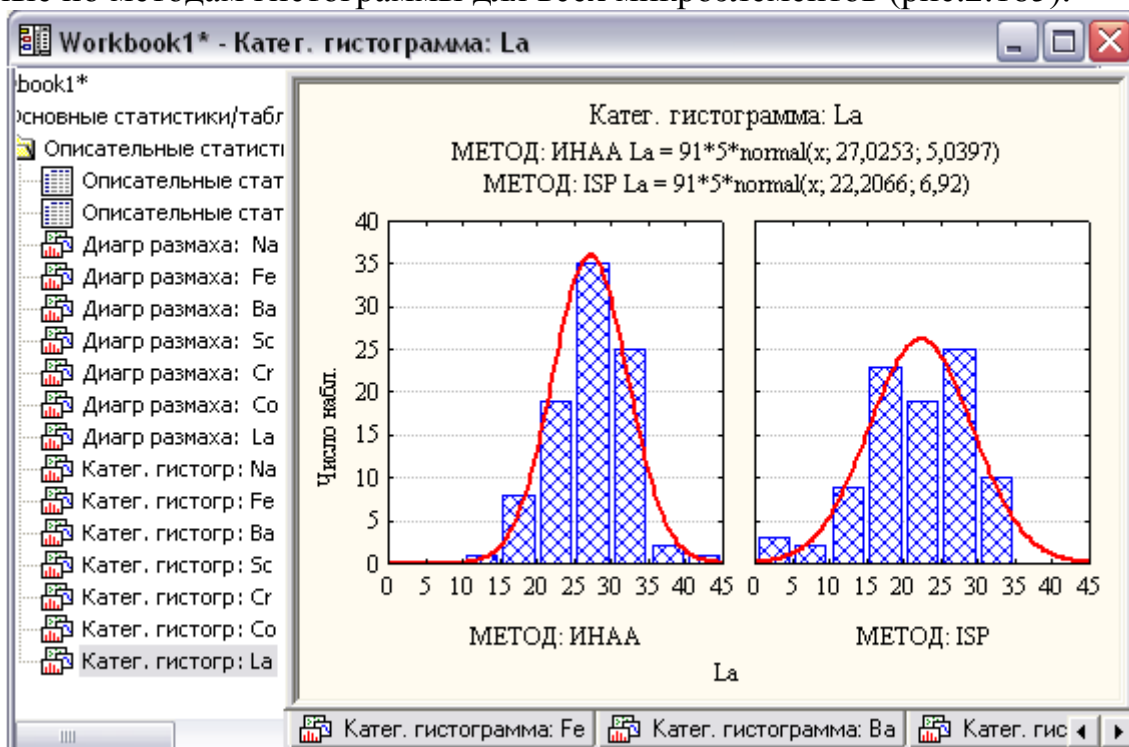



Рис. 2.185. Составная по методам гистограмма для микроэлемент La

При этом слева от картинки отражен столбцом протокол результатов рассмотренного (шаги 1-12) сценария анализа. Достаточно лишь щелкнуть мышкой по соответствующему ярлыку для просмотра интересующих нас результатов, например **Описательные статистики** для просмотра

таблицы числовых характеристик (рис.2.186) или  Диаграмма размаха: La для просмотра диаграммы размаха La (рис.2.187).

Workbook1* - Описательные статистики (ИНАА-ISP почва)

Переменная	Описательные статистики (ИНАА-ISP почва)			
	N набл.	Среднее	Медиана	Мода
Na _и	91	1,3330	1,3000	1,300000
Fe _и	91	3,1956	3,3000	3,200000
Ba _и	91	501,0989	480,0000	400,0000
Sc _и	91	12,0758	12,8000	12,80000
Cr _и	91	127,0055	117,4000	Множест.
Co _и	91	15,7945	15,3000	14,30000
La _и	91	27,0253	27,5000	Множест.
Na _г	91	1,7858	1,6973	1,990000
Fe _г	91	3,0415	3,0476	Множест.
Ba _г	91	406,6264	387,0000	380,0000
Sc _г	91	9,1802	9,0000	Множест.
Cr _г	91	117,5121	104,0000	80,00000
Co _г	91	18,6473	19,0000	Множест.
La _г	91	22,2066	24,0000	20,00000

Рис. 2.186. Просмотр таблицы числовых характеристик микроэлементов

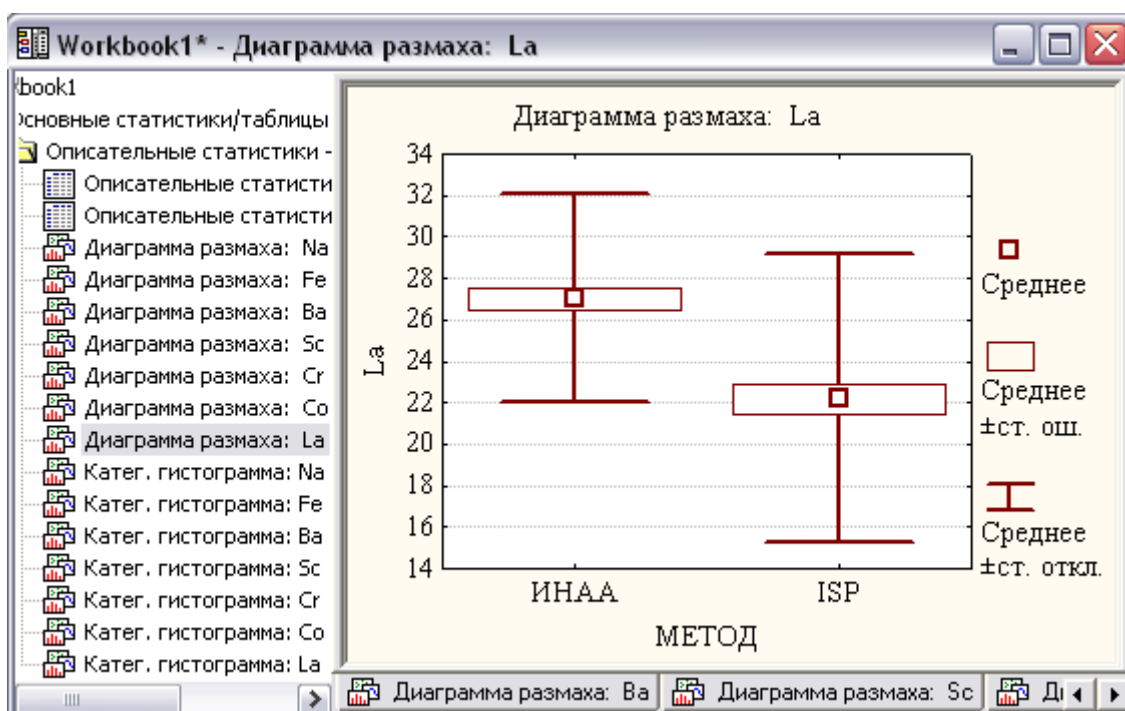



Рис. 2.187. Просмотр диаграммы размаха La по методам

Шаг 14. Появившееся на экране автономное окно (рис.2.190) нового макроса *МОС* содержит программу выполнения сценария анализа на языке **SVB**, доступную как для редактирования, так и для выполнения. Например, если в последней строке **True** поменять на **False**, то при последующем запуске макроса в таблице числовых характеристик микроэлементов (рис.2.186) исчезнет столбец *Мода*. Для запуска макроса следует нажать клавишу F5 или кнопку  **Выполнить Макрос** на панели Макрос. После чего выполнится сценарий запрограммированного выше статистического анализа (шаги 1-12).

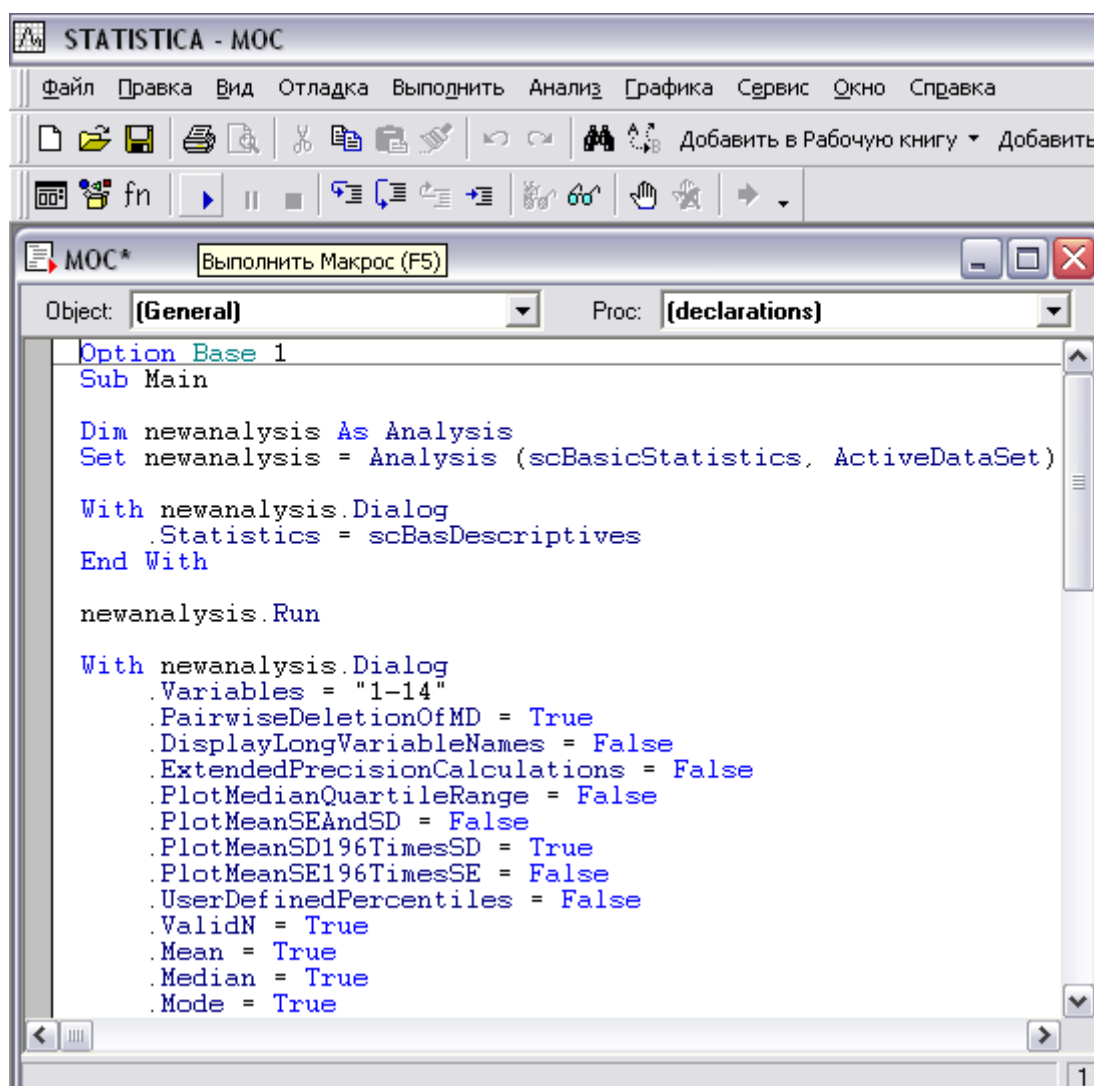


Рис. 2.190. Автономное окно нового макроса 

Особое внимание следует обратить на строку в центре автономного окна (рис.2.190): `Variables = " 1 - 14"`, содержащую номера столбцов вы-

бранных переменных базы данных (столбцы 1-14 на рис.2.1) для вычисления оценок числовых характеристик содержаний химических элементов. Эти параметры (" 1 - 14") также можно редактировать, что позволяет проигрывать сценарий МОС для другой базы данных.

Можно заметить, что редактированию доступны также таблицы (например, названия и шрифт числовых характеристик, а также шрифт и формат числовых ячеек) и рисунки (всевозможные корректировки как названий, так и оформления содержимого картинок) с помощью соответствующих средств системы.

На **этапе II** необходимо предварительно проверить гипотезу о нормальном распределении содержаний химических элементов (п.2.2.1) и применить критерий Стьюдента в случае нормально распределенных содержаний химических элементов и непараметрические критерии в остальных случаях (п.2.2.2).

Для реализации этого сценария анализа, включающего в себя несколько видов модулей, то есть создания мастер – макроса проверки гипотез (*ММПГ*) в активизированном пакете STATISTICA следует выполнить последовательность следующих действий:

Шаг 1. Для старта записи мастер – макроса запустите в головном меню модуль **Сервис** и выберите **Записать журнал Анализа (Мастер-макрос)...** в режиме **Макрос** (рис.2.191).

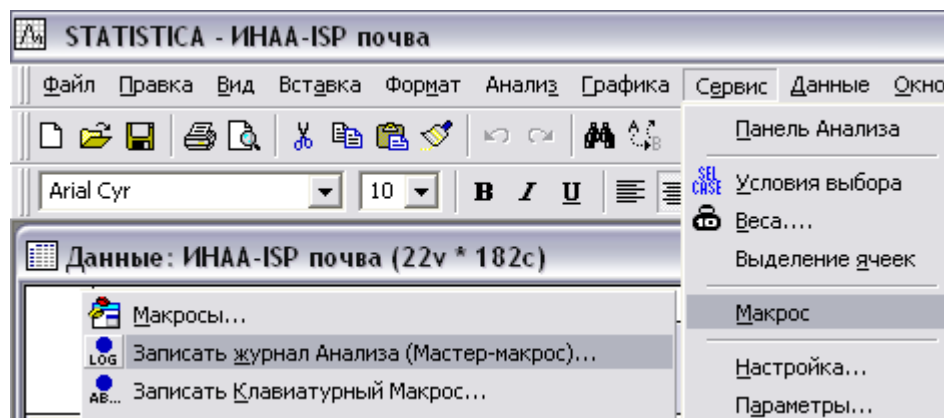
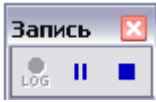





Рис. 2.191. Запуск записи мастер – макроса в головном меню

После старта записи мастер – макроса на экране появится панель

инструментов , позволяющая нажатием кнопки паузы  остановить временно запись, повторным нажатием кнопки паузы  продолжить запись и нажатием кнопки остановки  прекратить запись мастер – макроса.

Шаг 2. Для проверки гипотезы о нормальном распределении повторите шаги 1- 6 при построении рис.2.20 для одного микроэлемента.


Шаг 3. Продублируйте шаги 4 - 6 при построении рис.2.20 для 13-ти остальных микроэлементов (столбцы 1-14).

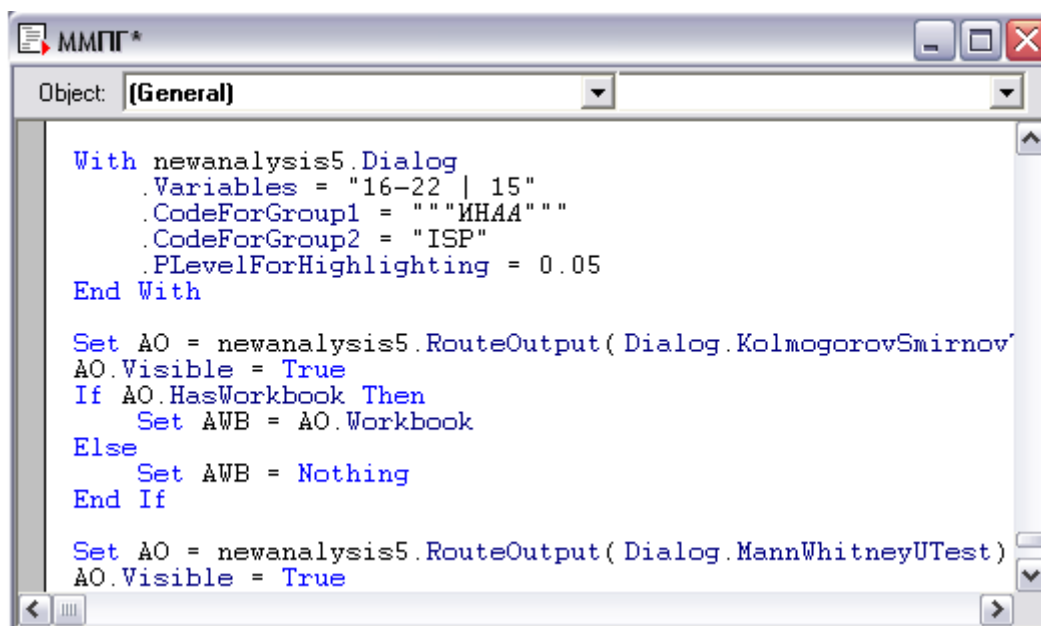
Шаг 4. Для визуального сравнения выборок (построения составных по методам гистограмм) повторите шаги 1- 6 при построении рис.2.17 для одного микроэлемента.

Шаг 5. Продублируйте шаг 4 при построении рис.2.17 для 6-ти остальных микроэлементов (столбцы 16-22).

Шаг 6. Для проведения аналитической оценки различия содержаний химических элементов по методам с помощью t - критерия Стьюдента повторите шаги 1- 5 при построении табл.2.2 .

Шаг 7. Для проведения аналитической оценки различия содержаний химических элементов по методам с помощью непараметрических критериев повторите шаги 1- 6 при построении табл.2.3-4 .

Шаг 8. Для прекращения записи макроса достаточно нажатия кнопки остановки , после чего появится окно для записи имени нового макроса *ММПП* (аналогично рис.2.167) с последующим выходом в автономное окно (рис.2.192).




```
MMPP*
Object: [General]

With newanalysis5.Dialog
    .Variables = "16-22 | 15"
    .CodeForGroup1 = ""ИНАА""
    .CodeForGroup2 = "ISP"
    .PLevelForHighlighting = 0.05
End With

Set AO = newanalysis5.RouteOutput( Dialog.KolmogorovSmirnov)
AO.Visible = True
If AO.HasWorkbook Then
    Set AWB = AO.Workbook
Else
    Set AWB = Nothing
End If

Set AO = newanalysis5.RouteOutput( Dialog.MannWhitneyUTest)
AO.Visible = True
```

Рис. 2.192. Автономное окно нового макроса *ММПП* (часть программы)

Для запуска макроса следует нажать клавишу F5 или кнопку  на панели Макрос. После чего выполнится сценарий запрограммированного выше статистического анализа, где можно посмотреть результаты применения всех 4-х модулей анализа, в частности, результаты проверки гипотезы о нормальном распределении микроэлементов

(рис.2.193) и результаты проведения аналитической оценки различия содержаний химических элементов по методам с помощью t -критерия Стьюдента (рис.2.194)

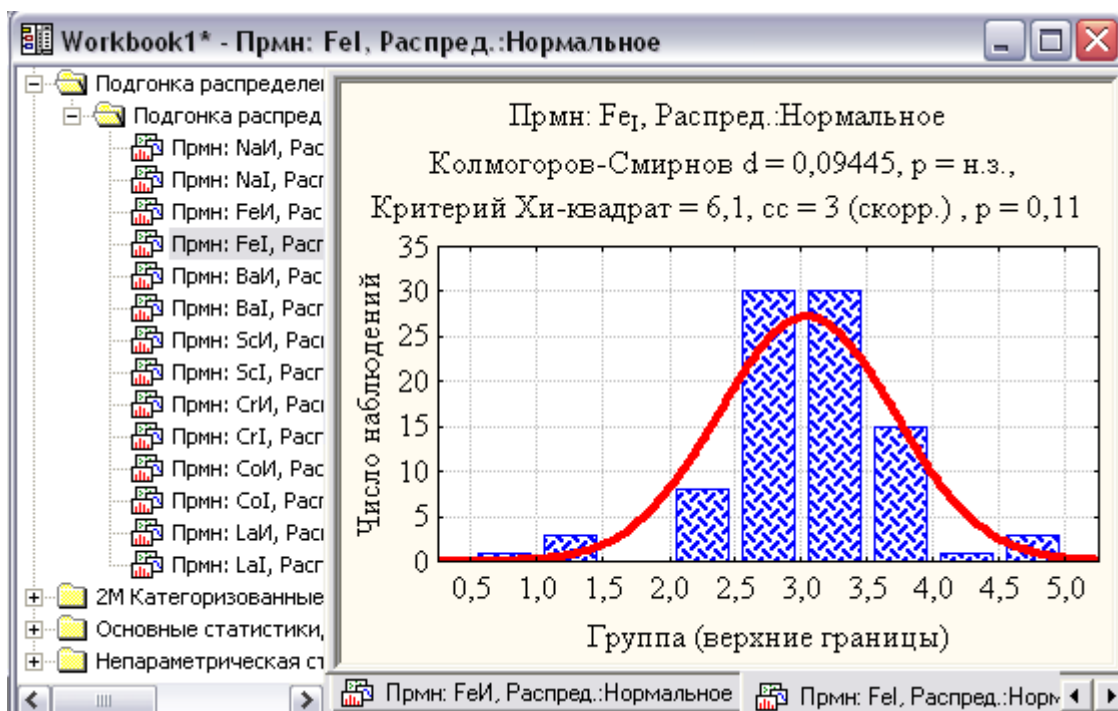


Рис. 2.193. ММПГ: результаты проверки гипотезы о нормальном распределении микроэлемента FeI

Перем	Т-критерии; Группир.: МЕТОД (ИНАА-ISP почва)				
	Среднее ИНАА	Среднее ISP	t-знач.	ст. св.	p
Na	1,33	1,79	-6,74	180	0,000000
Fe	3,20	3,04	1,55	180	0,123533
Ba	501,10	406,63	5,44	180	0,000000
Sc	12,08	9,18	7,33	180	0,000000
Cr	127,01	117,51	1,27	180	0,204925
Co	15,79	18,65	-2,42	180	0,016708
La	27,03	22,21	5,37	180	0,000000

Рис. 2.194. ММПГ: таблица результатов проведения аналитической оценки различия содержаний химических элементов по методам с помощью t -критерия Стьюдента

ПРИЛОЖЕНИЕ

П₁ Содержание микроэлементов (мг/кг) в почве территории Томского региона, полученное методами ИНАА (И) и ISP (I).

Таблица 2.17

Раздельная составляющая П₁

	Na _И	Fe _И	Ba _И	Sc _И	Cr _И	Co _И	La _И	Na _I	Fe _I	Ba _I	Sc _I	Cr _I	Co _I	La _I
Ед. Изм.	%	%	%	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %	%	%	%	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %
1	1,9	3,2	350	13,2	157,5	15,9	32,4	2,16	3,16	370	10,2	144,5	21,1	28
2	2	3,5	530	13,8	118,6	16,7	34,2	2,27	3,22	350	9,1	161,2	21,6	26,2
3	1,9	3,7	460	14,5	106,2	16,1	32,4	2,32	3,61	330	10	127,7	23	24,9
4	1,9	3,9	480	15,3	137,6	17,5	35,3	2,54	3,82	380	13,4	136,7	22,6	33,1
5	2,1	3,7	370	14,4	111,6	16,3	34,9	2,46	3,26	320	9,7	136,4	22,1	27,5
6	1,8	3	450	12,4	184,2	13,5	30,8	1,99	2,78	420	10,2	213,5	16,8	26,2
7	1,8	3,2	360	12,8	163,5	18,5	30,1	2,47	2,99	380	11,2	73,5	20,9	27
8	1,9	3,6	340	14,4	173,5	18,1	30,8	2,58	3,28	370	11,3	76,7	24	27,8
9	1,4	3,7	470	15,2	127,3	17,1	29,6	2,62	3,62	350	11	173	22,8	26,4
10	2	3,1	430	12,3	354,2	14,6	30,5	2,23	2,36	430	9,8	103,5	17,2	27,6
11	2	3,4	370	13,7	202	19,2	30,2	2,8	3,03	430	11,8	62,9	24,8	28,1
12	1,6	3,8	280	15,3	137,1	18,2	28,6	2,76	4	400	13,5	125	23,1	27,3
13	1,5	3,8	450	14,9	140,8	18,9	30,6	2,79	3,88	390	13,4	79,2	22,8	26,6
14	1,8	4,2	730	16,7	155,6	20	30,1	2,92	4,18	400	14,3	148,4	24	31,1
15	1,9	4,1	480	16,8	306,5	21,5	34,5	2,21	2,54	410	10	64,9	19,9	30,8
16	1,9	3,1	470	13,3	224,3	17,6	30,8	2,45	2,58	390	10,9	64,8	19,8	29,7
17	1,6	3,7	290	14,6	144,8	17,6	30,7	2,96	3,72	390	13,6	128	23	28,9
18	1,3	3,8	350	15,2	126,3	17,1	29,1	2,6	3,85	380	14	115,4	23,3	31,4
19	1,2	4	360	15,9	113,6	18,3	27,9	2,78	3,97	380	14	116,3	23,1	30,2
20	1,2	2,2	350	8,3	85,7	9,1	28,1	2,66	4,57	450	14	210	17	16
21	0,9	1,6	350	5,5	81,9	6,6	16,2	1,63	3,1	410	10	120	12	16
22	1,2	1,5	400	6	46,5	6,8	18,9	1,57	2,03	400	6	80	7	9
23	1,4	2	570	8,7	55,9	9,2	28,6	1,62	2,25	370	7	220	9	13
24	0,95	2	570	7,1	179,4	33,3	22,4	1,59	2,61	420	8	120	11	14
25	1,1	1,6	610	5,8	93,6	12	20,6	1,54	2,99	570	8	360	61	13
26	1,2	2,4	260	8,2	159,9	14,3	24,4	1,39	2,06	600	6	190	15	8
27	0,82	2,1	400	5,8	105,6	6,3	20,5	1,47	3,07	360	7	150	8	13
28	0,88	2,3	550	8,3	95,9	3,7	21,8	1,63	3,28	380	9	110	11	16

Продолжение табл. 2.17

	Na _{II}	Fe _{II}	Ba _{II}	Sc _{II}	Cr _{II}	Co _{II}	La _{II}	Na _I	Fe _I	Ba _I	Sc _I	Cr _I	Co _I	La _I
29	0,85	1,8	380	6,8	93,6	6,6	19,8	1,42	2,81	280	9	140	8	15
30	1,6	3,2	590	15,2	120,2	13,6	26	1,56	2,83	331	8	122	18	17
31	1,8	3,1	550	10,5	114,6	13,5	23	1,95	2,59	387	8	88	14	20
32	1,8	2,8	480	11,1	111	13,8	24,7	1,99	2,42	351	7	88	15	18
33	1,6	3,2	650	13,2	124,1	14,9	24,6	1,75	2,97	489	8	119	18	21
34	1,4	3,3	740	11,2	129,8	14,5	27,7	1,57	2,70	416	7	110	16	20
35	1,1	3,6	970	10,2	300,7	64,4	22,8	1,64	3,50	883	8	351	65	21
36	0,9	3,2	750	9,5	181,4	18,9	17,4	1,25	2,88	543	7	182	20	20
37	1,4	3	450	11,9	102,1	16,4	28,1	2,14	2,92	410	9	94	22	26
38	1,1	3,2	530	9,2	271,1	40	20,6	1,43	2,57	562	6	194	38	15
39	1,5	3,6	570	14,1	150,8	15,3	25,6	1,57	2,96	332	8	136	18	24
40	1,4	3	360	12,2	114,2	14,3	28,1	1,86	2,67	383	9	90	17	25
41	1,3	3,7	430	14,8	119,3	14,3	27,6	1,88	3,34	337	9	104	18	25
42	1,2	3,2	570	12,4	105,3	15	26,2	1,70	2,92	387	9	86	19	27
43	1,7	3,6	590	14,3	120,7	16,4	31,9	1,89	3,05	245	7	100	20	17
44	1,4	3,8	460	14,4	111,1	15,7	29,7	1,85	3,45	320	9	93	21	20
45	1,3	3,3	478	12,7	124,9	18,7	23,3	1,78	3,01	396	10	104	21	25
46	1,3	3,5	500	12,8	140,6	18,9	25,9	1,76	3,06	389	9	110	21	22
47	1,3	3,6	410	13,7	135,5	16,8	24,1	1,86	3,18	382	10	122	20	25
48	1,2	3,8	510	12,8	127,2	15,4	26,6	1,85	3,18	323	7	93	18	18
49	1,1	3,3	750	11,2	125,9	14,9	23,3	1,55	2,86	384	8	104	15	20
50	1,3	3,4	630	12	133,2	14,1	27	2,02	3,11	352	9	105	17	21
51	1,1	2,8	700	10,1	97,7	11,7	19	1,30	2,16	349	6	71	13	17
52	1,2	4	390	15,2	104,5	17,5	29,9	1,74	3,69	356	9	105	23	25
53	1	3,1	700	13,3	97,2	16,4	25,5	1,60	2,93	379	9	80	20	24
54	1	3,6	770	13,6	97,9	17,6	27,4	2,19	3,33	433	11	88	23	31
55	1,2	3,9	510	14,7	135,8	15,9	32,3	1,54	3,74	425	11	139	22	31
56	0,9	3,4	560	13,5	100,7	13,1	25,2	1,59	3,27	388	10	103	20	28
57	0,8	3,3	642	12,9	91,6	15,7	28,5	1,21	3,05	444	10	80	19	28
58	0,9	3,3	400	14,5	92,2	16,7	26,6	1,60	3,25	429	10	88	22	30
59	1	3,7	550	14,2	103,5	17	25,2	1,58	3,71	378	12	95	22	32
60	0,9	3,1	420	12,5	93,3	13,7	29,4	1,43	3,28	398	10	101	21	29
61	1	3,4	590	14,3	100,7	15	29,3	1,28	3,11	367	10	79	20	30
62	0,9	3,2	670	10,8	102,5	14,3	27,4	1,28	2,77	449	8	79	16	23
63	0,9	3,2	360	12	131,2	14,3	28,8	1,67	3,21	460	9	137	20	28

Окончаниетабл. 2.17

	Na _{II}	Fe _{II}	Ba _{II}	Sc _{II}	Cr _{II}	Co _{II}	La _{II}	Na _I	Fe _I	Ba _I	Sc _I	Cr _I	Co _I	La _I
64	1,1	3	390	12,2	114,5	15	30,6	1,73	2,85	440	9	103	19	28
65	1,3	2,9	520	11,9	104,2	13,1	26,9	1,41	2,65	399	8	95	17	24
66	1,1	3,5	510	13,2	132,4	16,2	41,1	1,37	3,25	408	10	128	22	30
67	1,3	3,1	590	10,5	131	13,6	26,2	1,60	2,72	408	7	119	15	21
68	0,8	2,6	580	8,4	102,4	14,1	19,8	1,11	2,38	502	6	90	16	21
69	1	3,9	570	14,8	87,9	18	36,1	1,73	4,56	487	13	97	26	31
70	0,8	3,6	680	10,5	131,8	14,6	27,5	1,27	3,48	556	7	126	18	25
71	1,1	4,4	490	16,2	99,7	18,9	34,8	1,77	4,51	436	13	98	27	33
72	0,8	3,8	740	13,3	127	17,8	25,6	1,53	3,93	529	11	105	24	28
73	0,7	2,7	560	8,9	138,2	15,7	24,5	1,27	2,98	451	8	110	18	24
74	0,6	2,6	410	9,2	156,4	12	21,4	0,88	2,18	339	6	159	14	18
75	0,8	3,7	400	12,8	104,6	16,7	26,6	1,54	3,70	528	10	115	22	26
76	1	2	440	7,3	117,4	8,7	21,3	1,02	1,42	353	4	89	8	13
77	1,4	3,6	520	14	106,8	11,1	33,5	1,46	2,89	379	8	107	19	22
78	0,5	3,1	890	8,6	151,4	15,1	20,6	1,07	3,00	782	7	153	17	20
79	0,8	4	680	11,6	127,9	14,9	26,4	1,25	3,20	518	7	93	16	19
80	0,9	4,8	380	12,2	82,9	13,9	24,3	1,17	3,57	361	7	65	14	17
81	1,7	3,2	450	12,4	82,6	17,2	30,1	1,52	2,97	380	10	100	11	18
82	2	3,3	490	13,5	92,1	17	31,1	1,83	3,07	350	11	90	13	20
83	2,1	3,7	400	15	81,8	17,8	33,6	1,99	3,79	340	14	100	14	19
84	1,9	1,9	520	6,9	68,7	8,1	18,7	1,06	1,05	380	4	90	3	2
85	1,3	1,6	410	5,4	52,2	7,5	14	1,2	1,12	340	4	70	3	4
86	1,7	1,8	400	6,3	56,8	8,2	15,8	1,07	0,99	310	3	70	3	3
87	1,5	2,8	420	13	179	13	22	1,78	2,55	400	9	150	9	15
88	2,1	3	310	12,1	84,7	14,2	30,7	1,99	2,69	310	9	120	10	14
89	2,3	3,1	390	12,8	88,3	13,9	31,7	1,96	2,72	290	8	80	10	16
90	1,7	3,4	370	14,2	84,1	15,1	33,4	1,81	3,31	300	10	80	11	17
91	1,5	3,3	400	15,5	175	16,6	27,5	1,82	2,92	360	10	100	11	18

Таблица 2.18

Категоризированная составляющая П₁

	Метод	Na	Fe	Ba	Sc	Cr	Co	La
Ед. изм.		%	%	%	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %	10 ⁻⁴ %
1	ИНАА	1,9	3,2	350	13,2	157,5	15,9	32,4
2	ИНАА	2	3,5	530	13,8	118,6	16,7	34,2
3	ИНАА	1,9	3,7	460	14,5	106,2	16,1	32,4
4	ИНАА	1,9	3,9	480	15,3	137,6	17,5	35,3
5	ИНАА	2,1	3,7	370	14,4	111,6	16,3	34,9
6	ИНАА	1,8	3	450	12,4	184,2	13,5	30,8
...
86	ИНАА	1,7	1,8	400	6,3	56,8	8,2	15,8
87	ИНАА	1,5	2,8	420	13	179	13	22
88	ИНАА	2,1	3	310	12,1	84,7	14,2	30,7
89	ИНАА	2,3	3,1	390	12,8	88,3	13,9	31,7
90	ИНАА	1,7	3,4	370	14,2	84,1	15,1	33,4
91	ИНАА	1,5	3,3	400	15,5	175	16,6	27,5
92	ISP	2,16	3,16	370	10,2	144,5	21,1	28
93	ISP	2,27	3,22	350	9,1	161,2	21,6	26,2
94	ISP	2,32	3,61	330	10	127,7	23	24,9
95	ISP	2,54	3,82	380	13,4	136,7	22,6	33,1
96	ISP	2,46	3,26	320	9,7	136,4	22,1	27,5
97	ISP	1,99	2,78	420	10,2	213,5	16,8	26,2
...
177	ISP	1,07	0,99	310	3	70	3	3
178	ISP	1,78	2,55	400	9	150	9	15
179	ISP	1,99	2,69	310	9	120	10	14
180	ISP	1,96	2,72	290	8	80	10	16
181	ISP	1,81	3,31	300	10	80	11	17
182	ISP	1,82	2,92	360	10	100	11	18