

Геологическая



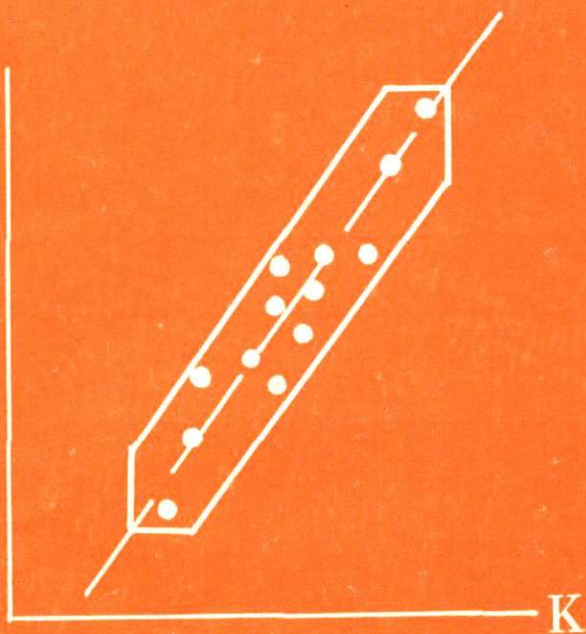
литература

«Наука»

Р.И.Лутков В.Н.Бондаренко

Математические модели зависимости геологических объектов

Rb





АКАДЕМИЯ НАУК СССР

МИНИСТЕРСТВО ГЕОЛОГИИ СССР

ИНСТИТУТ МИНЕРАЛОГИИ, ГЕОХИМИИ И КРИСТАЛЛОХИМИИ
РЕДКИХ ЭЛЕМЕНТОВ

Р.И.Лутков В.Н.Бондаренко

5154

Математические модели зависимости геологических объектов

Ответственный редактор
член-корреспондент АН СССР
Л. Н. ОВЧИННИКОВ



МОСКВА "НАУКА" 1989



Лутков Р.И., Бондаренко В.Н. Математические модели зависимости геологических объектов. — М.: Наука, 1989 — 120 с. — ISBN 5-02-003214-X

В монографии вводятся понятия поля регрессии, признака-аргумента и признака-функции; предлагаются методы расчета шкал двумерных диаграмм; описываются модели зависимости набора признаков в одном или нескольких геологических объектах. Разработан алгоритм программы исследования линейной зависимости на ЭВМ.

Расчитана на специалистов естественных наук.

Ил. 25. Табл. 33. Библиогр.: 25 назв.

Рецензенты: *Л.С. Бородин, Д.А. Родионов*

Lutkov R.I. and Bondarenko V.N. Mathematical models for the dependence of geological targets.

The monograph introduces notions and methods of defining the regression field, the indication-argument and indication-function; proposes methods for calculating scales of 2-dimentional diagrams; describes models for the relationship between sets of indications in one or several geological targets. An algoriphm for the program of studies of the linear dependence is proposed. The monograph is intended for specialists engaged in natural sciences.

Referces: *L.S. Borodin and D.A. Rodionov*

ВВЕДЕНИЕ

Геология как наука все меньше становится описательной. Большой объем фактов традиционных отраслей геологии и особенно из таких разделов, как геофизика и геохимия, требует новых методов обработки геологической информации. Разрабатываются новые виды анализов, внедряются в технику исследования новые приборы, привлекаются методы исследования других научных дисциплин. Особенно широко, как, впрочем, и в биологии и географии, используются математические методы исследования. В связи с тем что с позиции исследователя большинство геологических процессов имеет вероятностный характер, в качестве моделей для результатов деятельности этих процессов чаще всего используются вероятностные модели, а в качестве аппарата их изучения — аппарат математической статистики.

В научной и практической деятельности специалистов геологической отрасли традиционными и наиболее распространенными направлениями применения вероятностных моделей и аппарата математической статистики можно считать:

- а) подбор модели распределения значений признаков в геологических объектах и определение параметров распределения этих значений;
- б) сравнение параметров распределения значений признаков в геологических объектах и их группах;
- в) выделение статистически однородных групп геологических объектов;
- г) выявление зависимости значений геологических признаков и оценка ее вида.

Решению разнообразных задач геологии с применением математического аппарата в перечисленных выше аспектах посвящено значительное число исследований и публикаций. Необходимо отметить, что при решении широкого круга задач геологии в соответствии с общим развитием науки постоянно совершенствуется применение математических методов, расширяется их арсенал, ищутся и находятся новые аспекты более эффективного использования известных и малоизвестных методов. Все это определяет научно-технический прогресс как в геологической отрасли вообще, так и в применении математического аппарата в частности.

Среди перечисленных традиционных направлений применения методов математической статистики при решении задач в геологической практике существенный интерес представляет задача выявления зависимости значений двух признаков. Именно взаимоотношения значений признаков позво-

ляют исследователю получать информацию, которая не может быть получена другим способом.

В предлагаемой работе рассматриваются особенности применения к решению различных задач геологии наиболее простой, но широко распространенной модели зависимости значений геологических признаков – модели линейной зависимости.

Известно, что исследование прямолинейной зависимости в математической статистике осуществляется в рамках корреляционного и (или) регрессионного анализов. Корреляционный анализ предполагает изучение взаимозависимости двух равно случайных величин. В данной работе основное внимание уделено регрессионному анализу, предметом изучения которого является характер зависимости одного признака, квалифицируемого как функция, от другого, принимаемого за аргумент.

Применение методов выявления зависимости значений признаков при изучении состава колчеданных месторождений, особенностей взаимоотношений петрохимических характеристик, результатов определения абсолютного возраста и других проблем – все это позволило не только критически оценить эффективность применения традиционно используемых методов корреляционного и регрессионного анализов, но и осуществить поиски более простого и легко интерпретируемого решения этих задач. Немаловажным результатом этих поисков следует считать предложенную здесь методику, обеспечивающую простое и эффективное решение набора задач регрессионного анализа, реализация которого доступна широкому кругу геологов.

В связи с тем что в каждом конкретном случае интерпретация зависимости является прерогативой исследователя и, следовательно, конкретных рекомендаций по геологическому смыслу выявленной зависимости предлагаемая методика (как, впрочем, и любая другая) дать не может, то исследователю следует знать интерпретационные возможности применения методики, наиболее вероятные, допустимые и недопустимые области интерпретации. Этими факторами и обусловлена необходимость предварительных процедур определения однородности выборочной совокупности, характера распределения значений признаков, эмпирических значений параметров распределения и т.д. Следует указать, что в большинстве случаев при предварительном анализе выборочной совокупности используются широко известные статистические методы и критерии. В некоторых случаях предложены упрощенные методы, соответствующие общей задаче методики, но существенно уменьшающие затраты времени на вычислительные операции. В частности, можно отметить предложенные методы упрощенной оценки коэффициентов асимметрии и эксцесса распределения, необходимые при подборе модели распределения значений признаков.

Значительное внимание в работе уделено принципам и методам выбора осей для построения линии регрессии (аргумент–функция) и рационального изображения результатов анализа на двумерных диаграммах, без реализации которых невозможно эффективное применение методики. Указанными факторами определяется структура изложения материала в работе. Собственно регрессионный анализ (исследование зависимости двух признаков в одном объекте, а также обобщение методики на изучение характера зависимости двух признаков в наборе объектов или набора

признаков от одного аргумента в одном объекте) предворяется разделами, содержащими методику оценки параметров распределения признаков и выбора аргумента из изучаемых признаков.

Предлагаемая вниманию читателей монография отражает результаты многолетних исследований авторов по применению методов определения зависимости признаков на различных геологических объектах. Во многих случаях работа авторов осуществлялась в содружестве с другими исследователями и (или) с использованием их фактического материала. В главах I—III использован фактический материал авторов и А.А. Головина, основное содержание глав III и IV по применению методики регрессионного анализа написаны по результатам не только собственных, но и исследований, проведенных совместно с Л.Н. Овчинниковым и Л.С. Бородиным, которые любезно согласились на опубликование результатов этих исследований. Материал по машинному программированию предложенной методики регрессионного анализа подготовлен совместно с Л.Д. Караваевой. Своим соавторам в исследованиях, а также В.С. Воронову и другим коллегам за помощь при составлении программы "Lutar", за предоставленный фактический материал, за ценные советы и замечания авторы выражают свою искреннюю признательность и благодарность.

ИСПОЛЬЗУЕМЫЕ МАТЕМАТИЧЕСКИЕ СИМВОЛЫ

N	— число наблюдений (объем) обобщенной по всем объектам выборки (совокупности)
n	— число наблюдений (объем) частной выборки (совокупности)
\sum_1^n	— сумма всех единичных значений по совокупности объема n
ξ	— случайная величина, используемая в регрессионном анализе как независимая переменная (аргумент)
x_i	— эмпирические значения случайной величины ξ
\bar{x}	— среднее арифметическое — оценка математического ожидания случайной величины ξ
Δx_i	— разность между единичным наблюдением и оценкой среднего значения (\bar{x})
σ_x	— оценка стандартного отклонения от среднего значения случайной величины ξ
λ_x	— доверительный интервал оценки среднего случайной величины ξ для вероятности $(1 - q)$
$T_{\min x}$	— минимальное значение толерантного предела случайной величины ξ для вероятности $(1 - q)$
$T_{\max x}$	— максимальное значение толерантного предела случайной величины ξ для вероятности $(1 - q)$
V_x	— оценка коэффициента вариации для случайной величины ξ
K_x	— оценка коэффициента пропорциональной зависимости x от y
\tilde{K}_x	— оценка коэффициента регрессии x на y
$\tilde{\Delta}x_i$	— разность между единичным наблюдением и вычисленным из уравнения регрессии значением x

- $\tilde{\sigma}_x$ – оценка стандартного отклонения точек наблюдения от линии регрессии по оси абсцисс (x)
- $\tilde{\lambda}_x$ – доверительный интервал линии регрессии по оси абсцисс
- \tilde{T}_x – значение для толерантных пределов для отклонения точек наблюдения от линии регрессии по оси абсцисс, вычисленные для заданной вероятности $(1 - q)$
- D_x – оценка коэффициента диспропорциональности изменения значений x относительно значений y
- η – случайная величина, используемая в регрессионном анализе как зависимая переменная (функция)
- y_i – эмпирические значения случайной величины η
- \bar{y} – среднее арифметическое – оценка математического ожидания случайной величины η
- Δv_i – разность между единичным наблюдением и оценкой среднего значения (\bar{y})
- σ_y – оценка стандартного отклонения от среднего значения случайной величины η
- λ_y – доверительный интервал оценки среднего случайной величины η для вероятности $(1 - q)$
- $T_{\min y}$ – минимальное значение толерантного предела случайной величины η для вероятности $(1 - q)$
- $T_{\max y}$ – максимальное значение толерантного предела случайной величины η для вероятности $(1 - q)$
- V_y – оценка коэффициента вариации для случайной величины η
- K_y – оценка коэффициента пропорциональной зависимости y от x
- \tilde{K}_y – оценка коэффициента регрессии y на x
- $\tilde{\Delta} v_i$ – разность между единичным наблюдением и вычисленным из уравнения регрессии значением y
- $\tilde{\sigma}_y$ – оценка стандартного отклонения точек наблюдения от линии регрессии по оси ординат (y)
- $\tilde{\lambda}_y$ – доверительный интервал линии регрессии по оси ординат
- \tilde{T}_y – значение толерантных пределов для отклонения точек наблюдения от линии регрессии по оси ординат, вычисленные для заданной вероятности $(1 - q)$
- D_y – оценка коэффициента диспропорциональности изменения значений y относительно значений x
- $\Pi_{ас}$ – оценка показателя асимметрии для эмпирического распределения случайных величин ξ или η
- $\Pi_{экс}$ – оценка показателя эксцесса для эмпирического распределения случайных величин ξ или η
- $\tilde{\Pi}$ – оценка показателя эксцесса распределения точек наблюдения относительно линии регрессии
- t – значение случайной величины с распределением Стьюдента (критерий Стьюдента, t -критерий)
- $R_{x,y}$ – оценка парного коэффициента корреляции для случайных величин ξ и η

- q – вероятность ошибки I рода (уровень значимости)
- f – число степеней свободы
- \bar{x} – оценка генерального среднего случайной величины ξ , вычисленная по совокупности выборок
- \bar{y} – оценка генерального среднего случайной величины η , вычисленная по совокупности выборок
- \hat{K}_x – оценка коэффициента пропорциональности x по y , вычисленная по совокупности выборок
- \hat{K}_y – оценка коэффициента пропорциональности y по x , вычисленная по совокупности выборок

ОСОБЕННОСТИ ПРИМЕНЕНИЯ ТРАДИЦИОННЫХ МЕТОДОВ ОПРЕДЕЛЕНИЯ ЗАВИСИМОСТИ ПРИЗНАКОВ

При решении многих задач в геологии важное значение имеет изучение характера зависимости значений двух признаков в геологических объектах, моделями которых, как правило, являются случайные величины ξ и η . Такое изучение производится с целью выяснения: а) имеется ли какая-либо зависимость между значениями двух случайных величин? б) какова форма зависимости, в наилучшей степени отражающая связь значений этих величин, если она существует? в) можно ли по заданным значениям одной переменной предсказать значения другой и с какой точностью [17, 24, 25]?

Опыт применения математических моделей реальной зависимости значений признаков в геологических (и не только геологических) исследованиях однозначно показал, что в большинстве реальных ситуаций и условий конкретных задач наиболее рациональным является использование модели прямолинейной зависимости признаков. Причины этого положения определяются, во-первых, широким распространением зависимости признаков, близкой к линейной, во-вторых, сложностями процедурного характера, возникающими при использовании методов выявления и описания нелинейных моделей зависимости, в-третьих, простотой и надежностью как методов выявления и описания, так и, что особенно важно, интерпретации результатов применения модели прямолинейной зависимости.

Исследование зависимости признаков в геологической практике производится, как правило, в двумерных диаграммах $x - y$ (рис. 1). Из теоретически возможных вариантов расположения роя точек наблюдения в координатах двумерных диаграмм в подавляющем большинстве случаев характер зависимости признаков геологических объектов соответствует вариантам 2 и 13. Очевидна и причина этого вывода – специфика возможных значений признаков, характеризующих объекты геологических исследований и имеющих всегда положительные измерения (содержания химических элементов и минералов, запасы полезных компонентов, размеры рудных тел и др.).

Реже встречаются варианты 6 и 20, типичные для исследований с привлечением характеристик объектов с возможностью их отрицательных измерений (многие геофизические параметры: значения магнитных, гравиметрических и др. аномалий; расстояния объектов от какой-либо точки в ту или иную сторону; температуры, гипсометрические уровни и т.п.).

Остальные варианты из приведенных на рис. 1 хотя теоретически и возможны, при решении геологических задач практически не встречаются

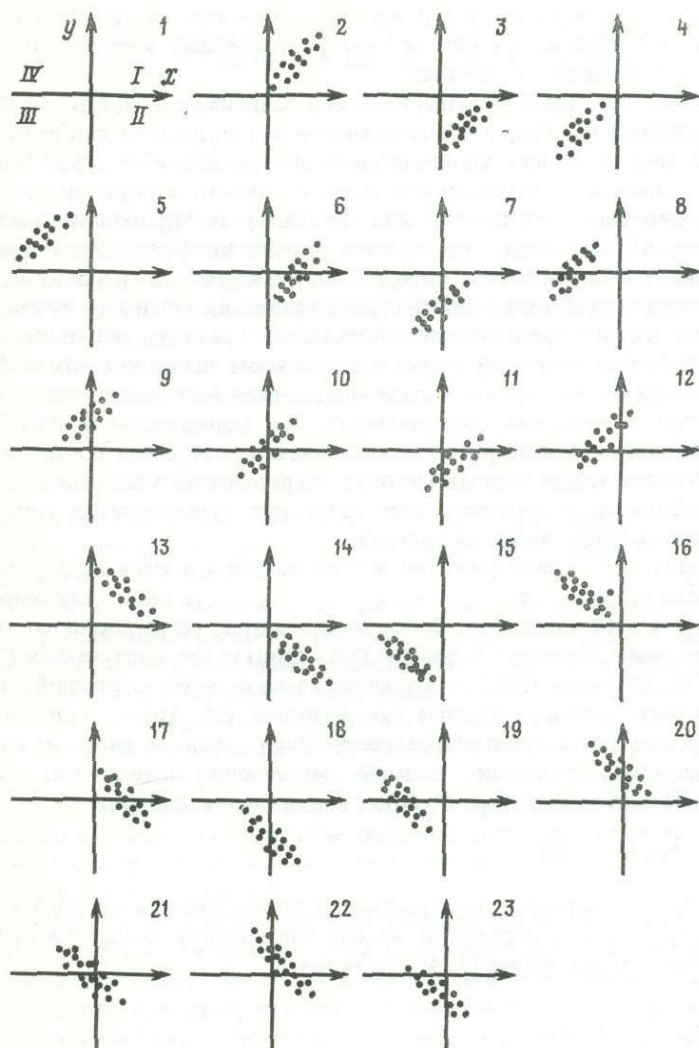


Рис. 1. Возможные варианты положения ориентированного роя точек на двумерной диаграмме

Исследование зависимостей этих вариантов соотношения признаков не имеет принципиальных отличий от вариантов 2 и 13. Можно отметить лишь нюансы в определении углов наклона линий регрессии к координатным осям (обязательной процедуры для определения роли признаков в их взаимоотношениях, состоящей из определения признака-аргумента и признака-функции). Заинтересованные исследователи могут определить эти углы на основании общих принципов выявления роли признаков, предложенных авторами на примере 2 и 13 вариантов.

Одной из наиболее простых (и, следовательно, наиболее удобных в прак-

гическом отношении) процедур изучения зависимости, близкой к линейной, является применение широко распространенных методов корреляционного и регрессионного анализов.

Известно, что задача о существовании зависимости между значениями двух случайных величин и о возможности использования для ее описания линейной модели вполне удовлетворительно решается методами корреляционного анализа – вычислением оценки парного коэффициента корреляции: рангового (Спирмена или Кендала) и "дробового выстрела" (Бломквиста). Результаты вычисления оценки коэффициента корреляции обеспечивают возможность определения: можно ли использовать линейную модель для описания реальной зависимости между признаками?

Более сложным представляется положение с реализацией методов определения значений одного признака при заданном значении другого. В то же время хорошо известно, что задачи такого рода весьма актуальны в прогнозно-металлогенических исследованиях, при формационном анализе рудных месторождений и в других не менее важных отраслях геологии. Однако применение традиционных методов регрессионного анализа к решению задач указанного содержания порождает ряд существенных трудностей. Рассмотрим это положение на примере.

Положим, что в распоряжении исследователя имеются выборочные совокупности $x_1, x_2, \dots, x_i, \dots, x_n$ и $y_1, y_2, \dots, y_i, \dots, y_n$ двух случайных величин ξ и η соответственно. Согласно процедуре регрессионного анализа, одна из величин, положим ξ , должна быть принята как неслучайная (закрепленная [1], аргумент [23]), а другая величина остается случайной и используется в регрессионном анализе как функция (η). Тогда, в соответствии с традиционным используемым в регрессионном анализе методом наименьших квадратов, уравнение прямой, отражающее зависимость значений функции от значений аргумента, будет иметь следующий вид:

$$\tilde{y}_i = \beta_0 + \beta_1 (x_i - \bar{x}),$$

где β_0 и β_1 – коэффициенты, отражающие положение прямой (β_0) и угла ее наклона (β_1) к оси абсцисс. В методе наименьших квадратов оценками коэффициентов уравнения [1, 4] являются

$$\beta_0 = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y};$$

т.е. оценка среднего случайной величины η (функции), и

$$\beta_1 = r_{x,y} \frac{\sigma_y}{\sigma_x},$$

где σ_y и σ_x – оценки стандартного отклонения η и ξ соответственно, $r_{x,y}$ – оценка коэффициента корреляции.

С учетом приведенных оценок эмпирическое уравнение регрессии η от ξ можно записать в следующем виде:

$$\tilde{y}_i = \bar{y} + r_{x,y} \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}).$$

Эмпирическое уравнение регрессии, отражающее зависимость значений x (теперь функция) от y (соответственно закреплённая величина, аргумент) имеет вид

$$\tilde{x}_i = \bar{x} + r_{x,y} \frac{\sigma_x}{\sigma_y} (y_i - \bar{y}).$$

Таким образом, зависимость¹ значений двух случайных величин ξ и η отображается двумя прямыми регрессии. Причем известно, что чем больше рассеяние точек наблюдения на двумерной диаграмме, тем больше угол между этими прямыми. Положение двух прямых регрессии полностью совпадает только при $r_{x,y} = \pm 1,000$. Если обозначить угол между двумя прямыми регрессии через γ , то зависимость его величины от оценок параметров уравнения регрессии можно записать следующим образом [1]:

$$\operatorname{tg} \gamma = \frac{1 - r_{x,y}^2}{r_{x,y}} \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \quad (I.1)$$

Из выражения (I.1) видно, что при $r_{x,y} = 0$ прямые регрессии (η от ξ и ξ от η) взаимно перпендикулярны, т.е. построение регрессионных прямых не имеет смысла.

Особенности применения метода наименьших квадратов при построении линий регрессии, приводящего к двум регрессионным прямым, во многих геологических задачах вызывают существенные затруднения как при практическом использовании результатов, так и при их интерпретации. Естественно, что метод наименьших квадратов в регрессионном анализе эффективен в тех случаях, когда одна из двух величин не является случайной. В геологии же подобные случаи редки, так как наиболее типичен случайный характер обоих признаков. Следовательно, эффективное применение регрессионного анализа в большинстве реальных ситуаций предполагает необходимость построения одной линии регрессии, одинаково точно отражающей зависимость как η от ξ , так и ξ от η .

Рассмотрим необходимость построения одной линии регрессии на примере анализа зависимости между двумя геологическими признаками, когда их роль во взаимоотношениях (аргумент или функция) неизвестна.

В табл. 1 дан исходный материал по содержаниям свинца и цинка в 28 элементарных ячейках геохимического ореола в разрезе № 18 Зырянского колчеданно-полиметаллического месторождения на Рудном Алтае [16].

Выборочная совокупность по содержаниям как свинца, так и цинка статистически однородна, распределение логарифмов содержаний металлов не противоречит нормальной модели распределения величин. Необходимые для регрессионного анализа параметры распределения содержаний металлов представлены в табл. 2.

Между содержаниями свинца и цинка в ореоле был определен коэффициент корреляции параметрическим методом $r_{x,y} = +0,778$ при критическом его значении $r_{0,05} (27) = +0,387$.

¹ Здесь и далее в тексте под определением "зависимость" подразумевается только линейная зависимость, так как другие формы зависимости в работе не рассматриваются.

Таблица 1

Содержания свинца и цинка (в г/т) в элементарных ячейках геохимического ореола Зырянского месторождения

Номер ячейки	Свинец	Цинк	Номер ячейки	Свинец	Цинк	Номер ячейки	Свинец	Цинк
1	54	139	11	106	650	21	515	1356
2	16	86	12	120	650	22	410	750
3	44	307	13	511	1225	23	159	447
4	189	239	14	536	826	24	27	90
5	521	1208	15	152	142	25	57	239
6	428	757	16	66	90	26	33	156
7	419	390	17	24	118	27	76	486
8	54	62	18	101	177	28	91	769
9	22	188	19	194	746			
10	19	80	20	302	1446			

Таблица 2

Параметры распределения содержаний металлов в геохимическом ореоле Зырянского месторождения

Статистические характеристики	Символы	Свинец	Цинк
Число наблюдений	N	28	28
Модель распределения		Логнормальная	Логнормальная
Среднее значение	$\bar{\lg x}, \bar{\lg y}$	2,034	2,511
Стандартное отклонение от среднего значения	$\sigma_{\lg x}, \sigma_{\lg y}$	0,493	0,429

Исходя из параметров распределения содержаний металлов и величины коэффициента корреляции между ними, были составлены уравнения регрессии по методу наименьших квадратов для свинца (в качестве аргумента в этом случае выступает цинк):

$$\lg \text{Pb} = 2,034 + 0,778 \frac{0,493}{0,429} (\lg \text{Zn} - 2,511). \quad (I.2)$$

и для цинка (в качестве аргумента принимаем свинец):

$$\lg \text{Zn} = 2,511 + 0,778 \frac{0,429}{0,493} (\lg \text{Pb} - 2,034). \quad (I.3)$$

Общий вид поля зависимости и линии регрессии свинца по цинку (А) и цинка по свинцу (Б) в геохимическом ореоле Зырянского месторождения можно видеть на рис. 2.

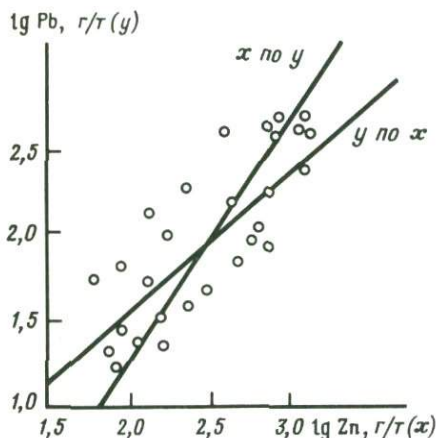


Рис. 2. Результат построения линий регрессии методом наименьших квадратов

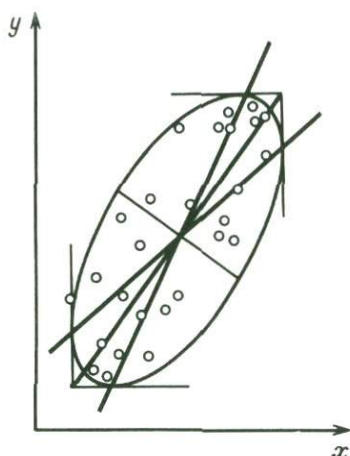


Рис. 3. Параметры эллипсоида рассеяния для роя точек, приведенного на рис. 2

На основании уравнения (1.2) определим содержание свинца в ореоле при заданном содержании цинка 100 г/г ($\lg Zn = 2,000$):

$$\lg Pb = 2,034 + 0,778 \frac{0,493}{0,429} (2,000 - 2,511) = 1,565$$

или, интерполируя из логарифмов, 37 г/г.

Принимая в качестве аргумента свинец, по уравнению (1.3) определим содержание цинка в ореоле, если содержание свинца там составляет 37 г/г ($\lg Pb = 1,565$):

$$\lg Zn = 2,511 + 0,778 \frac{0,429}{0,493} (1,565 - 2,034) = 2,201$$

или, интерполируя из логарифмов, 159 г/г.

Как видно из рассмотренного примера, построение уравнений и линий регрессии методом наименьших квадратов не является симметричным для обеих переменных: уравнение Pb по Zn не отражает зависимости Zn по Pb, тем более, что в метод наименьших квадратов не входит определение роли переменных в их взаимоотношениях — функция или аргумент.

Из приведенного рассмотрения очевидно, что в ситуациях, когда существует необходимость установления симметричной зависимости между обоими переменными, т.е. как выражения значений y через x , так и выражения x через y , применение регрессионного анализа с использованием метода наименьших квадратов нерационально. В таких случаях следует применять регрессионный анализ, основанный на использовании ортогональной регрессии.

Следуя рекомендациям работы [1], в целях наглядного представления соотношения регрессионных прямых, определяемых методом наименьших квадратов и прямой ортогональной регрессии, весьма продуктивным является использование геометрической характеристики концентрации

точек корреляционного поля (x_i, y_i) около своего "центра тяжести" — (\bar{x}, \bar{y}) . В случае двумерного нормального распределения величин ξ и η , модель которого принята в качестве наиболее распространенной и простой в геологических приложениях, наглядной характеристикой корреляционного поля является так называемый эллипс рассеяния. Уравнение его имеет следующий вид:

$$\frac{1}{1 - r_{x,y}^2} \left[\left(\frac{x_i - \bar{x}}{\sigma_x} \right)^2 - 2r_{x,y} \frac{x_i - \bar{x}}{\sigma_x} \frac{y_i - \bar{y}}{\sigma_y} + \left(\frac{y_i - \bar{y}}{\sigma_y} \right)^2 \right] = \chi_q^2(2), \quad (1.4)$$

где \bar{x} , \bar{y} , σ_x и σ_y — оценки параметров, $r_{x,y}$ — оценка парного коэффициента корреляции, а $\chi_q^2(2)$ — допустимое значение при уровне значимости q для случайной величины, распределенной как χ^2 с двумя степенями свободы. Точный (для двумерного нормального распределения ξ и η) смысл эллипса рассеяния состоит в том, что $100 \cdot (1 - q)$ процентов точек всей двумерной генеральной совокупности расположено внутри области, ограниченной эллипсом. Кроме того, форма и положение эллипса на двумерной диаграмме непосредственно связаны с регрессионными прямыми зависимости между ξ и η . Прямая ортогональной регрессии является большей из главных осей эллипса, а регрессионные прямые, построенные по методу наименьших квадратов, представляют собой диаметры этого эллипса, сопряженные с координатными осями (см. рис. 3).

Вслед за работой [1] механическую интерпретацию эллипса рассеяния и прямой ортогональной регрессии можно осуществить следующим образом. Припишем точкам (x_i, y_i) наблюдения внутри эллипса рассеяния единичные и равные массы, тогда ось, соответствующая наименьшему моменту инерции, совпадает с прямой ортогональной регрессии, проходящей через центр тяжести данной системы. Наибольшему моменту инерции рассматриваемой системы отвечает прямая, перпендикулярная линии ортогональной регрессии. Две указанные прямые и являются главными осями эллипса рассеяния.

Используя эллипс рассеяния, можно определять и положение линий регрессии, построенных по методу наименьших квадратов. Если построить касательные к эллипсу, параллельные осям координат, и соединить соответствующие точки касания, то мы получим собственно искомые прямые регрессии метода наименьших квадратов. По результатам построения прямых регрессии методом наименьших квадратов с помощью эллипса рассеяния может быть получено важное для дальнейшего изложения следствие. По определению (см. уравнение эллипса [1.4]) эллипс рассеяния ограничивает область точек, вероятность появления которых внутри эллипса равна $1 - q$ (0,95 в нашем случае). Следовательно, отрезки касательных, параллельные оси x , представляют собой толерантные пределы для эмпирических значений случайной величины η (y_i) при доверительной вероятности, равной $1 - q$. Аналогичные отрезки касательных, параллельные оси y , соответственно есть толерантные пределы для ξ . Отсюда следует, что сложные вычисления, связанные с решением эллипса рассеяния, можно значительно упростить, заменив их построением толерантных пределов для переменных при заданной доверительной вероятности $1 - q$.

Применение ортогональной регрессии требует предварительного исследо-

вания выборочных совокупностей, представляющих переменные. В первую очередь необходима оценка характера распределения значений случайных величин. Известно, что ортогональная регрессия наиболее эффективна при нормальном распределении значений переменных ξ и η . В связи с этим требуется предварительная проверка соответствия эмпирического распределения и модели нормального распределения. Из этого требования следует также необходимость определения однородности изучаемой двумерной совокупности. Кроме того, в отличие от регрессионного анализа с использованием метода наименьших квадратов, где нет необходимости в обосновании выбора аргумента, так как переменные играют эту роль поочередно, эффективность применения ортогональной регрессии предполагает необходимость однозначного выбора переменной, которую следует использовать в качестве аргумента.

В процедуре построения прямой ортогональной регрессии оценка коэффициента корреляции играет менее существенную роль, чем при использовании метода наименьших квадратов. В то же время проверка гипотезы о применимости модели прямолинейной зависимости для описания связи значений переменных, которая, как известно, основана на величине коэффициента корреляции, обуславливает необходимость его оценки.

Предполагаемая методика исследования зависимости между признаками геологических объектов с учетом особенностей конкретных задач геологии и применяемых статистических методов позволяет в ряде этапов существенно упростить процедуру без заметных потерь в обоснованности результатов. Опыт применения данной методики определил рациональную последовательность анализа выборочных совокупностей, обеспечивающую построение регрессионных зависимостей и их уверенную интерпретацию в свете конкретных задач геологических исследований. Исследование выборочных совокупностей предлагается осуществлять в следующем порядке.

1. Определение статистической однородности выборки по значениям анализируемых признаков и возможности представления их зависимости с помощью одной прямой регрессии.

2. Оценка характера и параметров распределения значений анализируемых переменных.

3. Установление пригодности линейной модели для описания реальной зависимости значений двух изучаемых переменных.

4. Определение взаимоотношений переменных, выбор аргумента и шкал значений признаков для построения графика зависимости признаков.

5. Регрессионный анализ и оценка уравнения ортогональной регрессии.

6. Оценка коэффициента парной корреляции значений переменных.

7. Графическое изображение зависимости значений двух переменных.

В соответствии с конкретной геологической задачей результаты исследования зависимости двух переменных могут быть обобщены на представление зависимости набора признаков от одного из них в одном объекте или на характеристику зависимости пары признаков в наборе геологических объектов.

ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ПЕРЕМЕННЫХ ПРИ ИССЛЕДОВАНИИ ИХ ЗАВИСИМОСТИ

Построение графика регрессионной зависимости является статистическим методом, базирующимся на процедуре оценки параметров. В связи с этим прежде чем осуществлять определение регрессионных зависимостей, необходимо выяснить структуру выборочных совокупностей и оценить параметры, описывающие ее положение в пространстве признаков.

Процедура предварительной оценки параметров включает: а) определение однородности выборочных совокупностей; б) подбор модели распределения значений признаков; в) вычисление оценок параметров распределения (среднего значения, стандартного отклонения, доверительного интервала для среднего значения); г) определение допустимых пределов в условиях соответствующей модели распределения и заданной доверительной вероятности; д) определение применимости линейной модели для описания реальной зависимости (вычисление предварительной оценки коэффициента корреляции и проверка гипотезы о ее значимости).

В большинстве ситуаций оценка параметров и проверка статистических стандартных гипотез осуществляются с помощью известных методов [1, 3, 4, 8, 15, 17, 19]. Для некоторых случаев авторами предложены упрощенные методы, соответствующие задаче исследования зависимости и обеспечивающие более простые способы реализации вычислительных операций и дающие сопоставимые результаты с традиционными методами.

II.1. Определение однородности выборочных совокупностей

Однородность исследуемой выборочной совокупности при применении модели линейной зависимости имеет весьма существенное значение. Известно [1, 5, 9, 17], что высокие значения эмпирического коэффициента корреляции и, как следствие, вывод о применимости линейной модели для описания реальной зависимости признаков могут быть обусловлены существенно разными причинами в условиях однородных и неоднородных совокупностей. Как следует из руководств по математической статистике, числовое значение эмпирического коэффициента корреляции отражает степень компактности точек наблюдения вокруг некоторой прямой. Оценка коэффициента корреляции равна ± 1 , если все точки лежат на одной прямой, и равна 0, если рой точек наблюдения на двумерной диаграмме образует круг. Степень отличия значения оценок коэффициента корреляции от нуля или единицы отвечает форме роя точек, промежуточной между прямой и кругом.

Если выборочная совокупность неоднородна, т.е. в ее составе присутствуют представители, как минимум, двух генеральных совокупностей с различными по величине двумерными средними, то значение оценки

коэффициента корреляции не отражает степень компактности точек вокруг прямой. В неоднородной совокупности, состоящей из двух подсовокупностей, значение оценки коэффициента корреляции отражает величину расстояния между двумерными математическими ожиданиями составляющих. Причем, чем больше это расстояние, тем больше значение выборочного коэффициента корреляции. В этом случае характер вытянутости роя в подсовокупностях влияет на значение оценки коэффициента корреляции только при небольших расстояниях между центрами подсовокупностей.

Существование двух факторов, определяющих высокие значения коэффициента корреляции, в первую очередь требует определения статистической однородности выборочной совокупности. Только при однородной совокупности может быть построена прямая регрессии с однозначной интерпретацией — она моделирует характер зависимости между значениями признаков. В случае неоднородных совокупностей подобная прямая не дает основания для приведенной выше интерпретации, так как эта прямая может соединять центры тяжести подсовокупностей или отражать не имеющую геологического смысла комбинацию влияния реальной зависимости в большей по объему подсовокупности и расстояния между ее центром и центром другой, меньшей по объему части выборки.

Таким образом, приведенное выше изложение свидетельствует о том, что прежде, чем приступить к построению прямой регрессии, необходимо определить однородность выборочной совокупности.

В своих исследованиях авторы применяют простой, но удобный в использовании метод проверки однородности выборочных совокупностей [11]. По этому методу строят гистограмму распределения значений признака и при ее многовершинности проверяют критерием Стюдента (t -критерием) гипотезу о равенстве максимумов и промежуточных минимумов гистограммы. Число интервалов гистограммы определяется по формуле Стреджесса:

$$H = 1 + 3,332 \lg N. \quad (П.1)$$

Шаг интервала гистограммы

$$C = \frac{x_{\max} - x_{\min}}{H}. \quad (П.2)$$

Значение t -критерия при разной величине максимумов:

а) максимумы встречаемости значений признака резко различны:

$$t = \frac{(n_1 - n_2)(1 - p)}{\sqrt{Np(1 - 2p) \cdot (1 + p)}}, \quad \text{при } p = \frac{n_1 + n_2}{2N},$$

где n_1 — абсолютная встречаемость в меньшем максимуме; n_2 — абсолютная встречаемость признака в соседнем промежуточном минимуме, p — средняя относительная встречаемость в меньшем максимуме и минимуме, N — общее число наблюдений в выборке;

б) максимумы встречаемости признака сопоставимы

$$t = \frac{n_1 + n_3 - 2n_2}{\sqrt{\frac{2N \cdot p(1-3p)(3-8p)}{(1-p)(1-2p)}}} \quad \text{при } p = \frac{n_1 + n_2 + n_3}{3N},$$

где n_1 и n_3 – абсолютная встречаемость признака в сравниваемых максимумах, n_2 – абсолютная встречаемость признака в промежуточном минимуме, p – средняя относительная встречаемость признака в сравниваемых максимумах и минимуме, N – общее число наблюдений в выборке.

Если эмпирическое (вычисленное) значение t -критерия больше допустимого при заданном уровне значимости и соответствующем числе степеней свободы, то нулевая гипотеза отвергается и выборка считается неоднородной.

В качестве примера рассмотрим определение однородности выборки рудных тел по размерам их по падению в одном из медных месторождений Закавказья (исходные данные представлены в графе 3 табл. 3). На рис. 4 приводится гистограмма распределения размеров рудных тел по падению, для которой число интервалов (согласно II.1):

$$H = 1 + 3,332 \lg 40 = 6,3 = 6 \text{ (с округлением);}$$

шаг интервала (согласно с II.2):

$$C = \frac{390 - 30}{6} = 60 \text{ м.}$$

Очевидно, что рассматриваемая выборочная совокупность состоит из двух подсовкупностей. Проверка t -критерием также отвергла гипотезу об однородности выборки:

$$p = \frac{10 + 4 + 9}{3 \cdot 40} = 0,19,$$

$$t = \frac{10 + 9 - 2 \cdot 4}{\frac{2 \cdot 40 \cdot 0,19 \cdot (1 - 3 \cdot 0,19)(3 - 8 \cdot 0,19)}{(1 - 0,19)(1 - 2 \cdot 0,19)}} = 2,51 > t_{0,05}(38) = 2,02.$$

По результатам проверки гистограммы на однородность исследуемой совокупности можно предположить, что на месторождении выделяются две группы рудных тел в зависимости от их размера по падению: 1) от 30 до 200 м, формировавшиеся в приповерхностных условиях. 2) от 200 до 390 м, формировавшиеся на большей глубине, чем первые.

Если взять выборку рудных тел в соседнем небольшом месторождении (точнее, в рудопроявлении), то из-за незначительного числа наблюдений (табл. 4) проверить однородность выборки рассмотренным методом невозможно.

Неоднородную совокупность при небольшом числе наблюдений ($n < 20$) на составляющие подсовкупности допустимо разделять ви-

Протяженность рудных тел по падению (ширина)
в одном из медных месторождений Закавказья

Номер п/п	Номер рудного тела	Ширина рудного тела, м x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
1	2	3	4	5	6
1	40	30	-177	31329	-5545233
2	17	37	-170	28900	-4913000
3	50	43	-164	26896	-4410944
4	5	49	-158	24964	-3944312
5	65	62	-145	21025	-3048625
6	10	71	-136	18496	-2515456
7	13	85	-122	14884	-1815848
8	14	95	-112	12544	-1404928
9	43	101	-106	11236	-1191016
10	1	108	-99	9801	-970299
11	75	114	-93	8649	-804357
12	38	122	-85	7225	-614125
13	11	130	-77	5929	-456533
14	70	133	-74	5476	-405224
15	64	136	-71	5041	-357911
16	30	141	-66	4356	-287496
17	45	148	-59	3481	-295379
18	25	157	-50	2500	-125000
19	7	172	-35	1225	-42875
20	44	191	-16	256	-4096
21	72	209	+2	4	8
22	31	223	16	256	4096
23	20	241	34	1156	39304
24	54	253	46	2116	97336
25	77	268	61	3721	226981
26	12	277	70	4900	343000
27	53	291	84	7056	529704
28	36	304	97	9409	912673
29	8	310	103	10609	1092727
30	56	315	108	11664	1259712
31	34	317	110	12100	1331000
32	35	320	113	12769	1442897
33	21	323	116	13456	1560896
34	23	328	121	14641	1771561
35	61	336	129	16641	2146689
36	27	342	135	18225	2460375
37	4	357	150	22500	3375000
38	78	370	163	26569	4330747
39	263	381	174	30276	5268024
40	3	390	183	33489	6128487
	Σ	8280	40301	495770	+1321560

зуальным анализом точек наблюдения на диаграмме после оценки коэффициента эксцесса распределения значений признака.

Оценка коэффициента эксцесса распределения значений признака определяется по формуле:

$$t_{\text{экс}} = \frac{d - Md}{\sqrt{Dd}},$$

где $d = \frac{1}{n\delta} \sum |x_i - \bar{x}|$, а величины Md и \sqrt{Dd} определяются по таблицам [3].

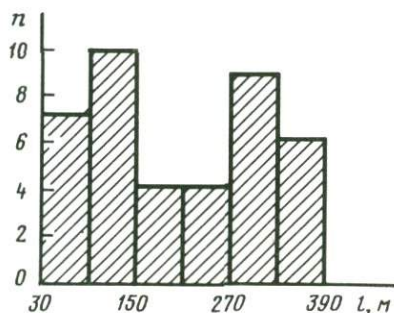


Рис. 4. Гистограмма распределения размеров рудных тел по падению одного из месторождений меди в Закавказье

Если вычисленное значение коэффициента эксцесса распределения превышает допустимое значение t -критерия при заданном уровне значимости и существующем числе степеней свободы, то считается, что распределение наблюдаемых значений признака не соответствует выбранной модели распределения величин.

Как показывает опыт статистической обработки геологической информации, при симметричном распределении значений признаков в исследуемых выборочных совокупностях оценка коэффициента эксцесса распределения значительно превышает допустимое значение в том случае, если эти выборочные совокупности не являются однородными, т.е. величина коэффициента эксцесса распределения может указывать на степень однородности совокупности (при условии симметричности распределения значений признаков).

Продемонстрируем данное положение на рассмотренном выше примере распределения размеров 40 рудных тел. В исследуемой выборке рудных тел по размерам их по падению вычисленная оценка коэффициента эксцесса распределения (см. табл. 3, графы 4, 5):

$$d = \frac{1}{40 \cdot 113} 4030 = 0,8916; \quad t = \frac{0,8916 - 0,8030}{0,0329} = 2,69 > t_{0,05}(38) = 2,02$$

превышает допустимое значение при заданных условиях t -критерия, что свидетельствует о несогласии наблюдаемого распределения размеров

Протяженность рудных тел по падению (ширина)
в одном из медных рудопроявлений Закавказья

Номер п/п	Номер рудного тела	Ширина рудного тела, м x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
1	3	83	-120	14400	-1728000
2	6	110	-93	8649	-804357
3	1	130	-73	5329	-389017
4	2	206	+3	9	27
5	7	281	78	6084	474552
6	4	301	98	9604	941192
7	5	310	107	11449	1225043
		1421	572	55524	-280560

рудных тел с нормальной моделью распределения величин, т.е. разделение общей совокупности рудных тел на две подсовокупности методом построения гистограммы распределения правомочно.

Используемые обычно [3] методы оценки коэффициента эксцесса распределения применимы к выборкам с числом наблюдений более 12 и требуют довольно трудоемких расчетов. В данной работе авторами предлагается упрощенный вариант расчета оценки коэффициента эксцесса распределения, не требующий сложных расчетов, таблиц, применимый к выборкам с любым (более 2) числом наблюдений, но дающий результаты, сопоставимые с результатами традиционных методов данного коэффициента.

Возьмем совокупность с функцией распределения [3, 20]:

$$\Phi_u = \frac{1}{\sqrt{2n'}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du$$

и определим, при каком отклонении от среднего значения будет группироваться половина выборки. Оказывается, что половина наблюдений группируется в интервале $\bar{x} \pm 0,5\sigma_x$. Принимая, что при симметричном и идеальном (по высоте кривой) распределении значений признаков оценка показателя эксцесса распределения должна быть равной 0,000, эту оценку можно определить из выражения:

$$P_{\text{экс}} = \frac{n(\bar{x} \pm 0,5\sigma_x)}{0,5N} - 1, \quad (\text{II.3})$$

где $n(\bar{x} \pm 0,5\sigma_x)$ — число наблюдений в интервале $\bar{x} \pm 0,5\sigma_x$, а N — общее число наблюдений в выборке. Если выборка "растянута" относительно среднего значения (уплощенная кривая распределения) или состоит из подсовокупностей (экстремально уплощенная кривая распределения), то оценка коэффициента эксцесса распределения > 0 , а если выборка однородна и "сжата" относительно среднего значения, то $P_{\text{экс}} < 0$. Как

можно понять из формулы (II.3), оценка показателя эксцесса распределения теоретически может меняться в пределах от +1,000 до -1,000. Для однородной совокупности, как показывает опыт, критическое значение показателя эксцесса распределения составляет +0,500. Если значение показателя эксцесса распределения $> +0,500$, то выборка, как правило, неоднородна и ее следует разделить на составляющие подсовкупности методом построения гистограммы распределения значений (при $n > 20$) либо визуальным анализом точек наблюдения на диаграмме (при $N < 20$).

Определим величину показателя эксцесса распределения по предлагаемой методике для выборки размеров 40 рудных тел (см. табл. 3).

Исходные данные: $N = 40$; $\bar{x} = \frac{8280}{40} = 207$; $\sigma_x = \frac{495770}{39} = 113$; число наблюдений в интервале значений $207 \pm 0,5 \cdot 113$ равно 7, отсюда

$$P_{\text{экс}} = \frac{7}{0,5 \cdot 40} - 1 = -0,650 < P_{\text{крит}} = -0,500.$$

Таким образом, для исследуемой неоднородной выборки величина показателя эксцесса распределения превышает допустимое значение, что сопоставимо с результатом традиционного метода определения коэффициента эксцесса.

Обратимся к выборке с небольшим числом наблюдений (см. табл. 4) и определим для ее значений показатель эксцесса распределения. Исходные данные:

$$N = 7; \bar{x} = \frac{1421}{7} = 203; \sigma_x = \frac{55524}{6} = 96;$$

число наблюдений в интервале значений $203 \pm 0,5 \cdot 96$ равно 1; отсюда:

$$P_{\text{экс}} = \frac{1}{0,5 \cdot 7} - 1 = -0,714 < P_{\text{крит}} = -0,500.$$

Как показывает оценка показателя эксцесса распределения, и эта выборка рудных тел неоднородна по их размерам, а состоит из двух групп рудных тел: а) с шириной рудных тел менее 200 м, б) с шириной рудных тел более 200 м.

Хотя показатель эксцесса распределения по принятой интерпретации математической статистики является показателем формы кривой распределения (уплощенной или заостренной), он может, как показал опыт его применения, указывать на неоднородность (экстремальную уплощенность) кривой плотности распределения, причем оценивать этот показатель можно по упрощенной методике.

II.2. Оценка характера и параметров распределения значений исследуемых переменных

Из предшествующего изложения известно, что эффективность исследования зависимости с применением метода построения прямой ортогональной регрессии во многом зависит от характера распределения значений изучаемых признаков. В связи с этим одним из основных этапов предварительного исследования выборочной совокупности является определение соответствия распределения эмпирических значений признаков модели двумерного нормального распределения. Сложность методов подбора модели распределения для многомерных (и даже двумерных) совокупностей и ориентации предлагаемой методики на простоту реализации определяют достаточность проверки гипотезы о модели распределения в одномерном варианте — для каждой из переменных самостоятельно.

Наиболее широко известным и распространенным способом определения пригодности модели нормального распределения для характеристики распределения эмпирических значений признака является метод, основанный на использовании коэффициентов асимметрии и эксцесса [3, 4, 20].

Кривая плотности нормального распределения, как известно из теории вероятностей и математической статистики, является симметричной относительно математического ожидания (среднего) генеральной совокупности. В связи с этим характеристика симметричной кривой эмпирического распределения при сопоставлении его с нормальной моделью в первую очередь предусматривает определение именно этого признака распределения значений в изучаемой выборке. Параметром, отражающим степень симметричности кривой распределения, является, как известно, третий центральный момент, или асимметрия распределения. Эмпирическое значение асимметрии распределения вычисляется по формуле

$$\gamma_3 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N\sigma^3},$$

где

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i,$$

а

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}.$$

Случайность—неслучайность выборочного значения асимметрии распределения определяется двумя способами. Наиболее широко распространен способ вычисления коэффициента асимметрии: $\gamma_3/\sigma_{\gamma_3}$, где σ_{γ_3} — стандартное отклонение оценки асимметрии, приближенное значение которого равно $\sqrt{\frac{6}{N}}$. Если абсолютное значение коэффициента асимметрии не превышает 3, то эмпирическое распределение совокупности призна-

ется симметричным. В противном случае распределение значений в выборочной совокупности следует считать существенно асимметричным.

Второй способ определения случайности полученного значения оценки асимметрии базируется на использовании точного распределения выборочного значения асимметрии и ее стандартного отклонения. Кривая эмпирического распределения значений признается симметричной, если величина

$$t_1 = \frac{\gamma_3}{\sqrt{D_{\gamma_3}}} \quad (\text{II.4})$$

не превысит допустимого значения случайной величины, распределенной нормально с параметрами 0 и 1 при заданном уровне значимости q . В приведенном выражении (II.4) D_{γ_3} (D_{g_1}) – дисперсия выборочной асимметрии, значение которой для соответствующего числа наблюдений определяется по таблице [3]. Если t_1 больше допустимого значения, то эмпирическое распределение следует считать существенно асимметричным.

Продемонстрируем рассмотренные выше методы на конкретном примере. В табл. 5 представлены выборочная совокупность содержаний серы в колчеданных месторождениях Урала с вкрапленным и прожилково-вкрапленным типом оруденения (содержания серы менее 20%) и необходимые расчеты для оценки параметров эмпирического распределения данной совокупности.

Оценка среднего значения содержания серы равна

$$\bar{x} = \frac{264,23}{20} = 13,21.$$

Выборочное значение стандартного отклонения для данного случая

$$\sigma = \sqrt{\frac{258,0363}{19}} = 3,69.$$

Оценка асимметрии распределения

$$\gamma_3 = \frac{-31,8995}{20 \cdot 50,2434} = -0,032.$$

Коэффициент асимметрии распределения в данном случае равен:

$$\frac{\gamma_3}{\sigma_{\gamma_3}} = \left| \frac{-0,032}{\sqrt{\frac{6}{20}}} \right| = \left| \frac{-0,032}{0,548} \right| = |-0,058|.$$

Значение коэффициента асимметрии эмпирического распределения оказалось по абсолютной величине менее 3. В связи с этим принимается предположение, что распределение исследуемой выборки симметрично относительно среднего и не противоречит нормальной модели распределения по основному признаку.

Вычисленное значение асимметрии распределения можно оценить и вто-

Результаты оценки параметров распределения содержания серы
в уральских колчеданных месторождениях

Номер п/п (<i>i</i>)	Содержание серы, % (x_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
1	6,00	-7,21	51,9841	-374,8054
2	7,73	-5,48	30,0304	-164,5666
3	8,51	-4,70	22,0900	-103,8230
4	9,46	-3,75	14,0625	-52,3438
5	10,00	-3,21	10,3041	-33,0706
6	10,94	-2,27	5,1529	-11,6971
7	12,66	-0,55	0,3025	-0,1664
8	13,03	-0,18	0,0324	-0,0058
9	13,17	-0,04	0,0016	-0,0001
10	13,55	+0,34	0,1156	+0,0393
11	13,93	0,72	0,5184	0,3732
12	14,00	0,79	0,6241	0,4930
13	14,08	0,87	0,7569	0,6585
14	14,20	0,99	0,9801	0,9703
15	14,71	1,50	2,2500	3,3750
16	14,96	1,75	3,0625	5,3594
17	15,83	2,62	6,8644	17,9847
18	18,80	4,59	21,0681	96,7026
19	19,67	6,46	41,7316	269,5861
20	20,00	6,79	46,1041	313,0468
$\sum_{i=1}^N$	$\Sigma = 264,23$		$\Sigma = 258,0363$	$\Sigma = -31,8995$

рым способом. Для нашего случая

$$t_1 = \left| \frac{-0,032}{0,4726} \right| = 0,067,$$

что значительно меньше допустимого значения t_1 для уровня значимости 0,05, равного 1,96.

Рассмотренные методы оценки симметричности выборочного распределения требуют вычисления эмпирического значения третьего момента, который в дальнейшей процедуре исследования зависимости не будет использован. Кроме того, асимметрия выборочного распределения из-за параметрического характера существенно зависит от конкретных значений признака. В частности, при рассмотренных способах вычисления на величину коэффициента асимметрии существенное влияние оказывают минимальные и максимальные значения признака, т.е. единичные значения выборки, далекие от наиболее часто встречающихся и, по сути дела, аномальные значения, определяют одну из основных характеристик распределения. Это положение, естественно, не может удовлетворять исследователя в условиях задачи определения характера зависимости признаков.

При разработке методики и ее опытной проверке установлено, что методы оценки асимметричности эмпирического распределения, принятые в процедурах подбора модели нормального распределения, могут быть заменены аналогичными по задачам, но более простыми по реализации методами.

Для процедуры построения зависимости предлагается следующий метод определения степени симметричности—асимметричности кривой плотности эмпирического распределения. Метод реализуется вычислением некоторого показателя асимметрии ($\Pi_{ас}$). Продемонстрируем предлагаемый метод на конкретном примере. Обратимся еще раз к табл. 5. Ранжируя содержания серы по возрастанию, можно заметить, что число месторождений с содержаниями серы, меньшими среднего арифметического выборки ($\bar{x} = 13,21\%$), равно 9, а с большими — 11. Из рассматриваемой выборочной совокупности явно следует, что чаще встречаются в изучаемом регионе месторождения с более высокими средними содержаниями серы, чем с низкими. Это отношение и вызывает асимметрию выборочного распределения. При равенстве встречаемости месторождений с низкими и высокими (по сравнению с \bar{x}) содержаниями серы их должно быть по 10. Только при таком соотношении выборочное распределение может быть признано симметричным, а показатель асимметрии распределения будет в точности равен 0. Если же число данных в выборке со значениями, меньшими среднего арифметического и большими его, не равно между собой, выборка может оказаться представителем генеральной совокупности с существенно асимметричным распределением, а степень асимметрии выборочного распределения может быть оценена с помощью показателя асимметрии, вычисляемого по формуле

$$\Pi_{ас} = 1 - \frac{n(x < \bar{x})}{0,5 \cdot N}, \quad (\text{П.5})$$

где $\Pi_{ас}$ — показатель асимметрии выборочного распределения; $n(x < \bar{x})$ — число наблюдений со значениями, меньшими \bar{x} ; N — общее число наблюдений в выборке.

Известно, что в случае симметричного распределения среднее совокупности (математическое ожидание) и медиана, разделяющая совокупность на две равные по числу наблюдений части, совпадают. Именно на этом положении основано выражение для приведенного выше показателя асимметрии: в случае точной симметричности кривой плотности выборочного распределения его значение будет равно 0. В качестве альтернативного, асимметричного распределения можно привести модель логнормального распределения, которая, как известно, находит широкое применение в геологических приложениях. В условиях указанной модели кривая плотности значений признака имеет четко выраженную положительную асимметрию. В то же время кривая плотности логарифмов этих значений имеет симметричное распределение и все свойства нормальной модели.

Из руководств по теории вероятности и математической статистике [4, 20] известно соотношение математического ожидания и медианы

в условиях логнормального распределения. Математическое ожидание случайной величины ξ (ее значений, а не логарифмов), т.е. $M_\xi = e^{\mu+1/2\sigma^2}$, а медиана $M_e = e^\mu$. Отсюда явно следует, что в условиях положительно асимметричного логнормального распределения медиана по величине всегда меньше, чем математическое ожидание. Можно утверждать, что это соотношение сохраняется и для выборочных совокупностей с положительной асимметричной кривой плотности распределения. Заменяя упоминавшиеся выше параметры генеральной совокупности (M_ξ и M_e) их оценками (\bar{x} и M_e), можно было бы определить степень асимметричности выборочного распределения их соотношения

$$1 - \frac{\bar{x}}{M_e}.$$

Нетрудно видеть, что в случае симметричности выборочного распределения это выражение будет равно 0, и соответственно его величина будет возрастать при повышении асимметричности кривой распределения. Однако это справедливо лишь в общем случае. Рассмотрим соотношение параметров при логнормальном распределении. Здесь

$$\frac{M_\xi}{M_e} = \frac{e^{\mu+1/2\sigma^2}}{e^\mu} = e^{\mu+1/2\sigma^2-\mu} = e^{1/2\sigma^2}.$$

Из этого явно следует, что повышение числового значения отношения M_ξ/M_e может быть обусловлено как возрастанием степени асимметричности кривой распределения и соответственно ростом разницы в значениях M_ξ и M_e , так и увеличением значения дисперсии при том же соотношении величин M_ξ и M_e .

В связи с этим, используя приведенный выше показатель асимметрии (Π_{ac}), который является практически непараметрическим (в нем косвенно используется лишь оценка математического ожидания), можно получить более устойчивые результаты. К тому же в рассматриваемой методике выявление точного соответствия выборочного распределения нормальной или логнормальной модели не является основной целью. В нашем случае необходимо лишь оценить, какое представление данных (собственно их значения или логарифмы) дает лучшее приближение зависимости между ними к модели прямолинейной зависимости.

Предлагаемый показатель асимметрии (Π_{ac}) не имеет известного распределения, поэтому жесткие границы его допустимых значений при заданном уровне значимости получены быть не могут. Пределы, различающие почти симметричные и существенно асимметричные выборочные распределения, могут быть предложены на основе некоторых предположений и опыта применения методики, приобретенном при изучении зависимости признаков в геологии.

Рассмотрим определение граничных значений показателя асимметрии для нормальной и логнормальной моделей на примере выборки содержания мышьяка из уральских колчеданных месторождений. Исходные данные и необходимые расчеты для определения показателя асимметрии представлены в табл. 6. Из этих данных определим среднеарифметическое

Содержания мышьяка в уральских колчеданных месторождениях

Номер п/п	Мышьяк, г/т x_i	$\lg x_i$	Номер п/п	Мышьяк, г/т x_i	$\lg x_i$	Номер п/п	Мышьяк, г/т x_i	$\lg x_i$
1	100	2,000	11	390	2,591	21	790	2,898
2	140	2,146	12	420	2,623	22	840	2,924
3	180	2,255	13	450	2,653	23	920	2,964
4	210	2,322	14	490	2,690	24	1000	3,000
5	240	2,380	15	520	2,716	25	1200	3,079
6	270	2,431	16	560	2,748	26	1300	3,114
7	300	2,477	17	590	2,771	27	1500	3,176
8	320	2,505	18	620	2,792	28	1800	3,255
9	340	2,523	19	670	2,826	29	2300	3,362
10	370	2,568	20	720	2,857	30	3000	3,477
						N		
						Σ	22550	82,132
						$i=1$		

значение содержаний мышьяка

$$\bar{x} = \frac{22550}{30} = 752 \text{ г/т}$$

и среднее значение логарифмов содержаний

$$\overline{\lg x} = \frac{82,132}{30} = 2,738.$$

По формуле (II.5) определим показатель асимметрии ($\Pi_{ас}$) для натуральных значений содержаний (предположив нормальную модель распределения содержаний):

$$\Pi_{ас1} = 1 - \frac{20}{0,5 \cdot 30} = -0,333$$

и для логарифмов значений (предположив логнормальную модель распределения):

$$\Pi_{ас2} = 1 - \frac{15}{0,5 \cdot 30} = 0,000.$$

Из результатов оценки показателя асимметрии очевидно, что для рассматриваемой выборки содержаний мышьяка лучше подходит логнормальная модель распределения. Из этих же результатов можно определить граничные значения показателя асимметрии распределения для нормальной и логнормальной моделей распределения.

Показатель асимметрии для симметричного логнормального распределения равен 0,000, в то же время для натуральных значений эта величина равна -0,333. В качестве граничного значения показателя асимметрии

при предлагается брать промежуточное значение между его величинами для логарифмов значений (при их симметричности распределения) и для натуральных значений

$$P_{ас\ крит} = \frac{P_{ас1} + P_{ас2}}{2}$$

или в числовом значении

$$P_{ас\ крит} = \frac{-0,333 + 0,000}{2} = -0,167.$$

Если в какой-либо исследуемой выборочной совокупности показатель асимметрии $P_{ас}$ для натуральных значений по абсолютной величине более $-0,167$, то для этой совокупности можно принять нормальную модель распределения, а если $P_{ас} < -0,167$, то следует проверять гипотезу о логнормальной модели.

Определение однородности и симметричности изучаемой совокупности значений признаков, зависимость между которыми является целью исследования, предполагает возможность оценки параметров их распределения, причем оценки достаточно надежной.

Представление значений признаков, характеризующих изучаемый геологический объект с помощью моделей случайных величин (ξ и η), предполагает необходимость характеристики их генеральных совокупностей. Из предположения о приблизительно нормальном распределении их значений (или логарифмов значений) следует, что совокупности случайных величин (ξ и η) полностью характеризуются двумя параметрами: математическим ожиданием (M_{ξ} и M_{η}) и дисперсией (σ_{ξ}^2 и σ_{η}^2).

Рассмотрим кратко условия оценивания указанных параметров, обеспечивающих достаточно высокие качества оценок в задаче исследования зависимости признаков.

Как известно [1, 3, 4, 17, 19, 20], к оценкам параметров распределения в математической статистике предъявляются следующие требования. Оценка должна быть несмещенной, эффективной и состоятельной. В соответствии с обсуждаемой в работе задачей содержание этих требований здесь не раскрывается, а читателям, заинтересованным в более полном знакомстве с этой проблемой, можно рекомендовать цитированные выше работы. Именно с этих позиций ниже и рассматриваются в кратком изложении процедуры оценки параметров распределения значений признаков.

Несмещенной, эффективной и состоятельной оценкой математического ожидания случайной величины (положим, ξ) является среднее арифметическое, вычисляемое с помощью выражения

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (II.6)$$

В условиях нормального распределения значений математическое ожидание является и центром распределения. Теми же свойствами обладает и математическое ожидание логарифмов значений при их нормаль-

ном распределении в условиях положительной асимметричности выборочного распределения значений. Среднее арифметическое значение логарифмов

$$\bar{\lg x} = \frac{1}{N} \sum_{i=1}^N \lg x_i \quad (\text{II.7})$$

также является несмещенной, эффективной и состоятельной оценкой математического ожидания величины $\lg \xi$. Следует помнить, что в условиях нормального распределения логарифмов (логнормальное распределение) среднее арифметическое, вычисленное по формуле (II.6), является несмещенной оценкой математического ожидания значений (M_{ξ}). Однако оно не только не эффективно, но вообще не является оценкой центра логнормального распределения. В связи с тем что в методе оценки зависимости с помощью ортогональной регрессии наибольшую существенную роль играет именно положение центра распределения двумерной случайной величины $\{\xi, \eta\}$, при асимметричном распределении значений признаков и симметричном – их логарифмов, в качестве необходимого параметра используется среднее арифметическое или значений или их логарифмов. Предпочтение отдается той оценке, которая вычислена по данным, совокупность которых обладает более симметричным распределением (значения или их логарифмы).

Дисперсия генеральной совокупности случайной величины (положим, ξ) может быть определена по выборочным наблюдениям с помощью выражения

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x}). \quad (\text{II.8})$$

Стандартное отклонение соответственно равно: $\sigma = \sqrt{\sigma^2}$. Как следует из выражения (II.8), приведенная оценка является несмещенной, эффективной и состоятельной в случае симметричного, близкого к нормальному распределению значений признака. В случаях, когда значения имеют асимметричное распределение, а их логарифмы – близкое к нормальному, то оценкой дисперсии с теми же свойствами является оценка, вычисленная с использованием выражения

$$\sigma_{\lg}^2 = \frac{1}{N-1} \sum_{i=1}^N (\lg x_i - \bar{\lg x})^2. \quad (\text{II.9})$$

Таким образом, в соответствии с характером распределения значений признака и соответственно с представлением информации о признаке в аппарате построения прямой ортогональной регрессии параметры распределения признаков определяются из выражений (II.6) и (II.8) – для значений или же из выражений (II.7) и (II.9) – для логарифмов значений.

В связи с тем что при исследовании зависимости между признаками и ее представлением в графическом изображении существенную роль играет поведение оценки среднего, еще одной необходимой операцией является построение доверительного интервала для среднего.

Положим, что для неизвестного математического ожидания μ случайной величины ξ по выборке вычислено значение оценки \bar{x} . Необходимо определить: насколько точно значение оценки \bar{x} отражает неизвестное значение μ ? Ответ на этот вопрос может быть получен (естественно, с вероятностных позиций) в результате реализации процедуры определения доверительного интервала для μ . Доверительный интервал μ определяется выражением

$$p(\bar{x} - \lambda \leq \mu \leq \bar{x} + \lambda) = 1 - q,$$

где

$$\lambda = t_{q/2} \cdot \frac{\sigma}{\sqrt{N}}.$$

В приведенном выражении $t_{q/2}$ — допустимое для вероятности $q/2$ значение случайной величины, распределенной по закону Стьюдента с $(n - 2)$ степенями свободы, а \bar{x} , σ и μ известны из предшествующего изложения. Вероятностный смысл доверительного интервала заключается в следующем: со сколь угодно большой заданной вероятностью $(1 - q)$ математическое ожидание генеральной совокупности μ находится в интервале $x \pm \lambda$.

Таким образом, в соответствии с результатами определения однородности и симметричности распределения выборочной совокупности реализация процедуры оценки параметров дает возможность осуществить построение линии регрессии, отражающей характер зависимости между признаками.

В главе I было указано, что в качестве основной модели для описания реальной зависимости между значениями признаков в данной работе рассматривается модель прямолинейной регрессии. Там же обосновано применение метода построения прямой ортогональной регрессии, как наиболее точно соответствующей характеру зависимости признаков в рассматриваемом классе задач геологии.

Как известно [1], построение прямой ортогональной регрессии основано на использовании понятия об эллипсоиде рассеяния, длинная ось которого при заданной доверительной вероятности $(1 - q)$ представляет собой линию ортогональной регрессии. Выше уже отмечалось, что построение эллипса рассеяния является достаточно трудоемкой процедурой, поэтому здесь предлагается заменить процедуру построения эллипса рассеяния методом определения толерантных пределов для той же доверительной вероятности. В этом случае прямая ортогональной регрессии соединяет точки пересечения ограничительных линий, соответствующих минимальным и максимальным значениям толерантных пределов для переменных (x и y). Рассмотрим метод построения толерантных пределов для переменных на примере одной из них (положим, ξ). Необходимо оговорить, что проверка однородности и симметричности выборочной совокупности дает основание предполагать приблизительно нормальное распределение генеральной совокупности анализируемой выборки (для значений или логарифмов значений признаков в соответствии с результатами проверки симметричности выборочного распределения).

Положим, что $\xi_1, \xi_2, \dots, \xi_n$ — взаимно независимые и одинаково

распределенные случайные величины с функцией распределения $F(x)$. Толерантными пределами распределения $F(x)$, соответствующими вероятности $P = 1 - q$, называют такие две функции $T_1(\xi_1, \xi_2, \dots, \xi_n)$ и $T_2(\xi_1, \xi_2, \dots, \xi_n)$, для которых событие $\{F(T_2) - F(T_1) \geq P\}$ практически достоверно. Причем вероятность этого события γ ($\gamma \approx 1$) называется коэффициентом доверия.

Вероятностный смысл толерантных пределов можно представить как критические значения случайной величины ξ , соответствующие вероятностям $q/2$ и $1 - q/2$.

В нашем случае с предположительно нормальным распределением и неизвестными параметрами (μ и σ^2) в качестве толерантных пределов обычно используются функции вида $\mu \pm K \cdot \sigma$. В данном случае

$$\mu = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{и} \quad \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2},$$

а толерантный множитель K является решением уравнения

$$P \left\{ \Phi \left(\frac{\bar{x} + K \cdot \sigma - \mu}{\sigma} \right) - \Phi \left(\frac{\bar{x} - K \cdot \sigma - \mu}{\sigma} \right) \geq P \right\} = \gamma.$$

Приведенное уравнение означает, что с вероятностью γ , близкой к единице, внутри интервала $\bar{x} \pm K \sigma$ находится не менее, чем доля $P = 1 - q$ генеральной совокупности. Толерантный множитель K не зависит от μ и σ и представляет собой функцию трех переменных: N , γ и $P = 1 - q$.

Вычисление точного значения K является весьма трудоемкой операцией. В связи с этим авторы используют значения толерантного множителя из работы [3].

В данной таблице число наблюдений в анализируемой выборке ограничено ($n \leq 50$), поэтому при необходимости вычисления толерантных пределов для выборок большого объема определение толерантного множителя следует осуществлять с использованием выражения

$$K \approx \sqrt{\frac{N-1}{\chi^2(100\gamma, N-1)}} \cdot \Psi \left(\frac{1+p}{2} \right) \cdot \left\{ 1 + \frac{1}{2N} - \frac{2[\Psi \left(\frac{1+p}{2} \right)^2 - 3]}{24N^2} \right\}.$$

В приведенном выражении $\Psi \left(\frac{1+p}{2} \right)$ — функция, обратная функции нормального распределения для вероятности $\left(\frac{1+p}{2} \right)$ (см. табл. 1.3 в [20]),

а $\chi^2(100\gamma, N-1)$ — процентная точка распределения χ^2 для вероятности 100γ и $(N-1)$ степеней свободы (см. табл. 2.2а в [3]).

Аналогичным способом с использованием выражения $\bar{x} \pm K \cdot \sigma_x$ могут быть построены толерантные пределы и для второй переменной y . Естественно, что в этом случае \bar{x} следует заменить на \bar{y} и подставить соответствующее значение стандартного отклонения для данной переменной (σ_y). Толерантный множитель K используется тот же.

В качестве толерантного множителя можно также использовать критерий Стьюдента, значения которого для разных уровней значимости и степеней

свободы приводятся почти во всех статистических справочниках и которым авторы пользуются в своих исследованиях. В этом случае толерантные пределы распределения значений признаков имеют следующие выражения:

$$T_x = \bar{x} \pm t \sigma_x; \quad T_y = \bar{y} \pm t \sigma_y.$$

Подводя итоги изложенному, авторы предлагают следующую процедуру определения оценок параметров распределения значений признаков.

1. Оценка среднего

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

2. Оценка показателя асимметрии распределения

$$П_{асx} = 1 - \frac{n(y_i < \bar{y})}{0,5N}; \quad П_{асy} = 1 - \frac{n(y_i < \bar{y})}{0,5N}.$$

(Если для какой-либо переменной $П_{ас} < -0,167$, то для нее рассматривается логнормальная модель распределения значений, а процедура определения оценок параметров рассмотрена ниже).

3. Оценка стандартного отклонения

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}; \quad \sigma_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}.$$

4. Оценка показателя эксцесса распределения

$$П_{эксx} = \frac{n(\bar{x} \pm 0,5 \sigma_x)}{0,5N} - 1; \quad П_{эксy} = \frac{n(\bar{y} \pm 0,5 \sigma_y)}{0,5N} - 1.$$

(Если для какой-либо переменной $П_{экс} < -0,500$, то выборку проверяют на однородность (см. раздел II.1.), и оценки параметров определяют для составляющих однородных совокупностей переменных).

5. Оценка доверительного интервала для среднего значения

$$\lambda_x = \pm t \frac{\sigma_x}{\sqrt{N}}; \quad \lambda_y = \pm t \frac{\sigma_y}{\sqrt{N}}.$$

6. Оценка доверительных толерантных пределов значений при заданной доверительной вероятности $(1-q)$ и существующих степенях свободы:

а) минимальные пределы значений

$$T_{x_1} = \bar{x} - t \sigma_x; \quad T_{y_1} = \bar{y} - t \sigma_y;$$

б) максимальные пределы значений

$$T_{x_2} = \bar{x} + t \sigma_x; \quad T_{y_2} = \bar{y} + t \sigma_y.$$

При $П_{ас}$ для натуральных значений какой-либо переменной (например, x), меньше $-0,167$, проверяют гипотезу о логнормальной модели распределения значений и вычисляют оценки для логарифмов значений переменной.

1. Оценка среднего значения для логарифмов

$$\bar{\lg x} = \frac{1}{N} \sum_{i=1}^N \lg x_i.$$

2. Оценка показателя асимметрии распределения

$$P_{\text{ас} \lg x} = 1 - \frac{n(\lg x_i < \bar{\lg x})}{0,5 N}.$$

3. Оценка стандартного отклонения

$$\sigma_{\lg x} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\lg x_i - \bar{\lg x})^2}.$$

4. Оценка показателя эксцесса распределения

$$P_{\text{экс} \lg x} = \frac{n(\bar{\lg x} \pm 0,5 \sigma_{\lg x})}{0,5 N} - 1.$$

5. Оценка доверительного интервала для среднего значения

$$\lambda_{\lg x} = \pm t \frac{\sigma_{\lg x}}{\sqrt{N}}.$$

6. Оценка толерантных пределов значений при заданной доверительной вероятности $(1-q)$ и существующих степенях свободы

а) минимальные пределы значений

$$T_{\lg x_1} = \bar{\lg x} - t \sigma_{\lg x};$$

б) максимальные пределы значений

$$T_{\lg x_2} = \bar{\lg x} + t \sigma_{\lg x}.$$

Определим по указанной процедуре оценки параметров распределения содержаний свинца и цинка в геохимическом ореоле (см. табл. 1), которые в дальнейшем будут использоваться при исследовании линейной зависимости содержаний этих металлов (глава III).

На основе данных табл. 1 определяем среднее значение x (содержания Zn) и y (содержания Pb)

$$\bar{x} = \frac{13819}{28} = 493,5; \quad \bar{y} = \frac{5246}{28} = 187,4$$

и показатель асимметрии, определив, что число наблюдений, меньших среднего значения, для x равно 17, а для y равно 19:

$$P_{\text{ас} x} = 1 - \frac{17}{0,5 \cdot 28} = -0,214; \quad P_{\text{ас} y} = 1 - \frac{18}{0,5 \cdot 28} = -0,286.$$

В связи с тем что показатели асимметрии распределения как для Zn, так и для Pb оказались выше критического $(-0,167)$, необходимо проверить

Содержания металлов и их логарифмы в элементарных ячейках геохимического ореола Зырянского месторождения

Номер ячейки	Цинк, г/т x	Свинец, г/т y	lg x	lg y	Номер ячейки	Цинк, г/т x	Свинец, г/т y	lg x	lg y
1	139	54	2,143	1,732	15	142	152	2,152	2,182
2	86	16	1,934	1,204	16	90	66	1,954	1,820
3	307	44	2,487	1,643	17	118	24	2,072	1,380
4	239	189	2,378	2,276	18	177	101	2,248	2,004
5	1208	521	3,082	2,717	19	746	194	2,873	2,288
6	757	428	2,879	2,631	20	1446	302	3,160	2,480
7	390	419	2,591	2,622	21	1356	515	3,132	2,712
8	62	54	1,792	1,732	22	750	410	2,875	2,613
9	188	22	2,274	1,342	23	447	159	2,650	2,201
10	80	19	1,903	1,279	24	90	27	1,954	1,431
11	650	106	2,813	2,025	25	239	57	2,378	1,756
12	650	120	2,813	2,079	26	156	33	2,193	1,519
13	1225	511	3,088	2,708	27	486	76	2,687	1,881
14	826	536	2,917	2,729	28	769	91	2,886	1,959

логнормальную модель распределения содержаний цинка и свинца. С этой целью натуральные значения содержаний металлов переводятся в логарифмы (табл. 7) и оценки параметров определяются для логарифмов содержаний.

1. Оценка среднего для логарифмов значений

$$\bar{\lg x} = \frac{70,308}{28} = 2,511; \quad \bar{\lg y} = \frac{56,945}{28} = 2,034.$$

2. Оценка показателя асимметрии распределения

$$П_{ас\lg x} = 1 - \frac{14}{0,5 \cdot 28} = 0,000; \quad П_{ас\lg y} = 1 - \frac{15}{0,5 \cdot 28} = -0,071.$$

3. Оценка стандартного отклонения для логарифмов значений

$$\sigma_{\lg x} = \frac{4,969}{27} = 0,429; \quad \sigma_{\lg y} = \frac{6,557}{27} = 0,493.$$

4. Оценка показателя эксцесса распределения

$$П_{экс\lg x} = \frac{7}{0,5 \cdot 28} - 1 \approx -0,500; \quad П_{экс\lg y} = \frac{9}{0,5 \cdot 28} - 1 \approx -0,357.$$

Гипотеза об однородности выборочной совокупности по содержаниям x (методику проверки см. в разделе II.1) подтвердилась:

$$t = 1,71 < t_{0,05}(26) = 2,06.$$

5. Оценка доверительного интервала для среднего значения:

$$\lambda_{lgx} = 2,06 \frac{0,429}{28} = \pm 0,167; \quad \lambda_{lgy} = 2,06 \frac{0,493}{28} = \pm 0,192.$$

6. Оценка толерантных пределов значений при доверительной вероятности 0,95:

а) минимальные

$$T_{lgx_1} = 2,511 - 2,06 \cdot 0,429 = 1,627,$$

$$T_{lgy_1} = 2,034 - 2,06 \cdot 0,493 = 1,018;$$

б) максимальные

$$T_{lgx_2} = 2,511 + 2,06 \cdot 0,429 = 3,395,$$

$$T_{lgy_2} = 2,034 + 2,06 \cdot 0,493 = 3,050.$$

Параметры распределения значений признаков можно вывести из параметров распределения их логарифмов. Антилогарифмированное среднее для логарифмов значений признака соответствует среднегеометрическому самих значений:

$$\text{anti lg } \overline{\lg x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}.$$

Следовательно, антилогарифмированное стандартное отклонение логарифмов представляет собой меру размаха значений признака при условии использования среднегеометрической оценки центра распределения

$$T_{x_{\min}} = \text{antilg } \overline{\lg x} : (\text{antilg } \sigma_{lgx})^2;$$

$$T_{x_{\max}} = \text{antilg } \overline{\lg x} (\text{antilg } \sigma_{lgx})^2.$$

Квадрат антилогарифма стандартного отклонения логарифмов показывает: **во сколько** раз минимальное допустимое значение признака меньше, а максимальное допустимое значение — больше среднегеометрического.

Сопоставим: при нормальной модели распределения значений стандартное отклонение показывает, **на сколько** минимальное или максимальное из допустимых значений отличается от средней оценки.

Рассмотрим для примера определение параметров распределения содержаний Pb в геохимическом ореоле (табл. 7). Параметры распределения логарифмов содержаний Pb:

$$\overline{\lg Pb} = 2,034; \quad \sigma_{lg Pb} = 0,493;$$

$$T_{lg Pb} = \overline{\lg Pb} \pm t \sigma_{lg Pb};$$

$$T_{lg Pb_{\min}} = 2,034 - 2,05 \cdot 0,493 = 1,023;$$

$$T_{lg Pb_{\max}} = 2,034 + 2,05 \cdot 0,493 = 3,045.$$

Антилогарифмированные значения параметров:

$$\text{antilg } \overline{\lg Pb} = 108 \text{ г/т}; \quad \text{antilg } \sigma_{lg Pb} = 3,11;$$

$$\text{antilg } T_{lg Pb_{\min}} = 10,6 \text{ г/т};$$

$$\text{antilg } T_{lg Pb_{\max}} = 1108 \text{ г/т}.$$

Оценки параметров распределения содержаний цинка и свинца в геохимическом ореоле Зыряновского месторождения

Номер п/п	Статистические оценки	Символы	Цинк x	Свинец y
1	Число наблюдений	N	28	28
2	Модель распределения		Логнор- мальная	Логнор- мальная
3	Среднее значение	$\bar{lg} x, \bar{lg} y$	2,511	2,034
4	Стандартное отклонение	$\sigma_{lgx}, \sigma_{lgy}$	0,429	0,493
5	Доверительный интервал для среднего значения	$\lambda_{lgx}, \lambda_{lgy}$	$\pm 0,167$	$\pm 0,192$
6	Показатель асимметрии	P_{ac}	0,000	-0,071
7	Показатель эксцесса	$P_{экс}$	-0,500	-0,357
8	Толерантные пределы значений при $(1-q) = 0,95$			
	а) минимальные	T_{lgx_1}, T_{lgy_1}	1,627	1,018
	б) максимальные	T_{lgx_2}, T_{lgy_2}	3,395	3,050

Эти же минимальные и максимальные допустимые значения содержаний Pb в геохимическом ореоле можно получить и другим путем:

$$T_{Pb_{min}} = 108 : (3,11)^{2,05} = 10,6 \text{ г/т,}$$

$$T_{Pb_{max}} = 108 \cdot (3,11)^{2,05} = 1108 \text{ г/т.}$$

Нетрудно видеть, что полученные минимальные и максимальные значения в $(3,11)^{2,05} = 10,24$ раза соответственно меньше и больше среднегеометрической оценки.

Такая интерпретация стандартного отклонения логарифмов значений обеспечивает устойчивое определение угла наклона линии регрессии при логнормальной модели распределения их значений и, следовательно, при отсутствии фиксированного значения начала координат диаграммы.

Процедура использования стандартного отклонения логарифмов значений признаков будет показана ниже, при исследовании зависимости логарифмов значений.

Несколько слов о представлении результатов процедуры определения параметров изучаемой выборки, предшествующей непосредственному исследованию зависимости. Для представления результатов оценки параметров распределения значений признаков и структуры выборки, а также для рационального построения регрессии с учетом распределения как значений признаков, так и логарифмов значений, предлагается стандартизация записей параметров выборочной совокупности. Из опыта применения рассматриваемой методики известно, что в дальнейших процедурах необходимы следующие параметры и сведения.

1. Объем выборки. 2. Оценки средних значений. 3. Оценки стандартного отклонения. 4. Оценки доверительных интервалов для средних значений.

5. Оценки показателя асимметрии распределения. 6. Оценки показателя эксцесса распределения. 7. Толерантные пределы значений.

Итоговым является указание, на основе какой именно информации рационально осуществлять построение линии регрессии: по значениям признаков или по логарифмам их значений. Кроме того, в стандартизированном бланке может даваться начало координат при использовании отрицательных значений признаков или логарифмов значений.

Примером подобной таблицы может служить табл. 8 результатов оценки параметров выборочной совокупности содержаний цинка и свинца в геохимическом ореоле Зырянковского колчеданно-полиметаллического месторождения (Рудный Алтай).

Макет стандартной таблицы оценки параметров распределения значений признаков дан также в приложении 4.

Проверка выборочной совокупности на однородность по значениям каждого признака, выбор модели распределения значений, оценка параметров распределения значений и выявление структуры выборки являются первым этапом исследования зависимости.

Глава III

РЕГРЕССИОННЫЙ АНАЛИЗ ПРИ ИССЛЕДОВАНИИ ЗАВИСИМОСТИ ЗНАЧЕНИЙ ДВУХ ПРИЗНАКОВ

Объекты геологии, формирование которых, как правило, завершилось миллионы и даже миллиарды лет назад, а в равной степени и процессы их образования, время проявления и длительность которых также недоступны непосредственному наблюдению, оставляют мало надежд на возможность прямого исследования этих объектов. В связи с этим основными методами изучения объектов геологии и процессов их образования является наиболее полное и эффективное использование косвенных данных и построение ретроспективных моделей процессов. В частности, в таких отраслях геологии, как петрология, геохимия, металлогения, теория рудных месторождений, определяющее значение имеют: установление динамики свойств объектов, в основном соответствующих признаков, в возрастном и латеральном аспектах, выявление взаимосвязности изменения концентраций химических элементов (минералов, руд и т.п.) в пределах одного объекта или в разных объектах (по возрасту, латерали, структурному положению, формационной принадлежности и т.д.).

Перечисленные задачи по содержательному смыслу в большинстве своем отвечают классу ситуаций, в которых эффективным способом выявления и представления зависимости является регрессионный анализ. Существует значительное число работ, посвященных как собственно методам регрессионного анализа, так и проблемам их применения к задачам геологии. В то же время нельзя сказать, что все возможности различных вариантов анализа зависимостей, их моделей с достаточной полнотой выявлены и определены применительно к значительному разнообразию условий и содержания задач.

III.1. Проверка применимости модели прямолинейной зависимости

Следующим этапом в исследовании зависимости после оценки параметров распределения анализируемой совокупности и ее симметричности и эксцесса является определение принципиальной возможности применения прямолинейной зависимости для описания реальных соотношений во взаимном изменении значений изучаемых признаков геологического объекта. Сформулированная задача относится по своему содержанию к классу задач корреляционного анализа, методы которого хорошо известны и широко применяются в геологических исследованиях различной направленности [1, 3, 4, 5, 8, 17].

Результаты корреляционного анализа в качестве составной части входят в такие современные методы, как анализ главных компонент, факторный анализ, которые заняли прочное место в процедурах обработки геологической информации. Однако в аспекте настоящей работы нас интересует только часть функций корреляционного анализа: определение применимости модели прямолинейной регрессии. В соответствии с этим из широкого набора методов определения пригодности модели линейной зависимости к реальным данным здесь рассматривается и предлагается к широкому применению при исследовании зависимости только один метод. Этот метод в главе I данной работы назван как коэффициент корреляции Бломквиста. В работах [4, 5] он фигурирует как коэффициент корреляции "дробового выстрела". Кроме того, в некоторых работах он определяется как "тетрахорический коэффициент корреляции Бломквиста".

Рассматриваемый и предлагаемый к применению метод отличается исключительной простотой вычисления значений коэффициента корреляции и реализации метода проверки его значимости. При этом он обладает не меньшей, а в некоторых распространенных ситуациях (наличие случайных выбросов или аномальных значений) существенно большей, чем широко используемый парный коэффициент корреляции, надежностью. Имеется в виду надежность установления "вытянутости" роя точек наблюдения и определение степени компактности его вокруг некоторой гипотетической прямой линии. Именно подобная структура выборочной совокупности отображает наличие зависимости между признаками в генеральной совокупности и возможность описания этой зависимости с помощью линейной модели.

Метод вычисления коэффициента корреляции "дробового выстрела" основан на анализе роя точек наблюдения, вынесенных на двумерную диаграмму в координатах x и y (эмпирические значения случайных величин ξ и η соответственно). Известно, что применение точечных диаграмм при вычислении коэффициента корреляции рационально при объеме выборочной совокупности, не превышающем 100 наблюдений. В случаях, когда объем выборки превышает 100 наблюдений и, возможно, значительно, то рекомендуется использовать табличный вариант метода. В связи с тем что в ситуациях, рассматриваемых в настоящей работе, объем выборок не превышает 100 наблюдений, табличный вариант вычисления коэффициента корреляции "дробового выстрела" не рассматривается. Заинтересованным в этом методе специалистам можно рекомендовать работы [4, 5], в кото-

рых детально обсуждается этот вариант метода и на конкретных примерах демонстрируется его реализация и применение.

Положим, что в нашем распоряжении имеется точечная диаграмма (рис. 5), отражающая взаимное положение точек наблюдения в координатах x и y . В данном случае x и y представляют собой логарифмы содержаний соответственно цинка и свинца в ореоле Зырянского месторождения. Проведем на данном графике медианы по каждому признаку, т.е. $Me(x)$ и $Me(y)$. Как уже отмечалось, медианой называется значение признака, делящее данную совокупность на две равные по объему части. Естественно, что выборочное значение медианы делит данную выборку на две части с равным числом наблюдения в них. В результате проведения медианных

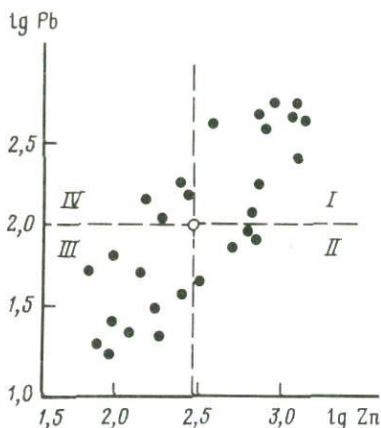


Рис. 5. Пример вычисления коэффициента корреляции методом "дробового выстрела"

прямых анализируемая точечная диаграмма оказалась разделенной на четыре квадранта (I, II, III, IV). Обозначим число наблюдений, находящихся в I и III квадрантах, через n_1 , а во II и IV – через n_2 (естественно, что $n_1 + n_2$ равно N – общему числу наблюдений в выборке). Тогда оценка коэффициента корреляции может быть вычислена с помощью выражения

$$r = \frac{n_1 - n_2}{n_1 + n_2} = \frac{n_1 - n_2}{N}. \quad (\text{III.1})$$

Вывод о применимости модели прямолинейной зависимости в данном случае базируется на результатах сопоставления полученной оценки с допустимым ее значением при заданном уровне значимости q . Выражение для допустимого значения оценки имеет вид

$$r_{\text{доп}} = \frac{\Psi(1-q)}{\sqrt{N^{-1}}},$$

где $\Psi(1-q)$ – значение функции, обратной функции нормального распределения с параметрами $(0, 1)$, для вероятности $(1-q)$; N – число наблюдений в изучаемой выборке. Если $r \leq r_{\text{доп}}$, то следует сделать вывод о невозможности применения линейной модели для описания зависимости

между признаками. Соотношение $r \geq r_{\text{доп}}$ определяет применимость линейной модели для описания реальной зависимости между значениями x и y по имеющимся в нашем распоряжении наблюдениям и при заданном уровне значимости 0,05.

Из приведенной двумерной диаграммы для рассматриваемого примера (см. рис. 5) определяем необходимые значения: $N = 28$, $n_1 = 22$; $n_2 = 6$. Используя приведенное выражение (III.1), определим оценку коэффициента корреляции между содержаниями цинка и свинца

$$r = \frac{22 - 6}{22 + 6} = \frac{16}{28} = 0,571.$$

Определим допустимое значение коэффициента корреляции при числе наблюдений $N = 28$ и двустороннем уровне значимости 0,05. В этом случае

$$r_{\text{доп}} = \frac{\Psi(1-0,025)}{28} = \frac{1,96}{4,6} = 0,426.$$

В связи с тем что в рассматриваемом примере $r > r_{\text{доп}}$, линейная модель может быть использована для описания реальной зависимости между признаками x ($\lg \text{Zn}$) и y ($\lg \text{Pb}$).

Из приведенной на рис. 5 точечной диаграммы следует еще один вывод. Точка пересечения медиан для признаков практически совпадает с положением двумерного среднего, координаты которого оценены по выборке (точка \bar{x} , \bar{y} – незалитый кружок). Этот факт свидетельствует о симметричности распределения выборочных значений по обоим признакам, что и было установлено при исследовании ранее этой характеристики изучаемой выборочной совокупности.

После определения, что по наблюдаемой выборочной совокупности, имеющейся в нашем распоряжении, зависимость значений признаков может быть описана с помощью линейной модели, предлагается выбрать наиболее рациональный способ представления этой зависимости с помощью линии регрессии.

III.2. Представление зависимости признаков с помощью модели пропорционального изменения их значений

Наиболее простой моделью регрессии, позволяющей эффективно использовать возможности ее применения в некоторых важных задачах, является модель пропорционального изменения значений изучаемых признаков.

Будем называть линию, соединяющую точки с координатами $(x_i = 0; y_i = 0)$ и $(x_i = \bar{x}; y_i = \bar{y})$, прямой пропорциональной зависимости. В данном случае имеется в виду положительная зависимость между признаками, которую часто определяют как "прямую пропорциональную зависимость". В случае отрицательного значения коэффициента корреляции – "обратная пропорциональная зависимость" – это требование справедливо для прямой, соединяющей точки с координатами $(x_i = 2\bar{x}; y_i = 0)$ и $(x_i = 0; y_i = 2\bar{y})$. В случае пропорциональной зависимости (прямой или обратной) оценка коэффициента корреляции по выборке любого объема равна ± 1 .

Модель пропорциональной зависимости полностью определяется одним параметром: коэффициентом пропорциональности

$$\hat{K}_x = \bar{x}/\bar{y} \text{ или } \hat{K}_y = \bar{y}/\bar{x}.$$

Геометрический смысл коэффициента пропорциональности достаточно просто продемонстрировать следующим образом. Если обозначить через α угол наклона линии пропорциональной зависимости к оси x , а через β соответственно к оси y , то очевидно, что $\operatorname{tg} \alpha = \hat{K}_y$, а $\operatorname{tg} \beta = \hat{K}_x$. Отсюда нетрудно видеть, что определение значений одной переменной по известному значению другой легко реализуется из выражений:

а) для прямой пропорциональной зависимости:

$$x_i = \hat{K}_x y_i; \quad y_i = \hat{K}_y x_i;$$

б) для обратной пропорциональной зависимости:

$$x_i = 2\bar{x} - \hat{K}_x y_i; \quad y_i = 2\bar{y} - \hat{K}_y x_i.$$

Из приведенных выражений явно следует, что направление (прямая или обратная) пропорциональной зависимости определяется предварительно — при вычислении оценки коэффициента корреляции (+1 или -1). Непосредственно из выражения коэффициента пропорциональности эта информация получена быть не может.

Необходимо отметить одну особенность применения модели пропорциональной зависимости. Приведенные выше выражения для модели пропорциональной зависимости применимы только для значений признаков, больше нуля. В то же время в практике геологических исследований могут встретиться ситуации, когда признаки имеют отрицательные значения (оценки некоторых геофизических параметров, абсолютные отметки геологических образований в рельефе и т.п.). При изучении петрохимических и геохимических данных часто встречаются значения признаков, меньше 1. Естественно, что в случаях, когда к линейной модели зависимости наиболее близки не сами содержания, а их логарифмы, то тоже возникает ситуация с отрицательными значениями признаков (логарифмами содержаний химических элементов). В данном случае приведенные выше параметры и методы приводят к абсурдным результатам — к абсурдным значениям коэффициентов пропорциональности.

Обратимся к выборке с прямой пропорциональной зависимостью $x_i = 17; 35; 52; 68; 78$ и $y_i = 0,136; 0,280; 0,416; 0,544; 0,624$. $N = 5$, $\bar{x} = 50$, $\bar{y} = 0,400$. Для этой выборки $\hat{K}_y = 0,400/50 = 0,008$; $\hat{K}_x = 50/0,400 = 125$. Рассмотрим изменение значений коэффициента пропорциональности при смещении выборки по осям координат таким образом, чтобы в выборке присутствовали отрицательные значения.

1. Выборочные значения получены смещением по оси ординат к началу координат на величину среднего $\bar{y} = 0,400$: $x_i = 17; 35; 52; 68; 75$. $\bar{x} = 50$; $y_i = -0,264; -0,120; 0,016; 0,144; 0,244$. $\bar{y} = 0,000$. При данных значениях признаков оценки коэффициентов пропорциональности получаются абсурдными $\hat{K}_y = 0/50 = 0$; $\hat{K}_x = 50/0 = \infty$.

Выборочные значения получены смещением по оси абсцисс влево на величину среднего $\bar{x} = 50$: $x_i = -33; -15; 2; 18; 28$. $\bar{x} = 0$. $y_i = 0,136;$

0,280; 0,416; 0,544; 0,624. $\bar{y} = 0,400$. И в этом случае оценки коэффициентов пропорциональности получились абсурдными

$$\hat{K}_y = \frac{400}{0} = \sim; \quad \hat{K}_x = \frac{0}{400} = 0,000.$$

Из приведенных примеров видно, что величина коэффициента пропорциональности зависит от точки отсчета значений. Приведенные ранее выражения коэффициентов пропорциональности при одной и той же зависимости, но при изменении точки отсчета могут принимать самые разные значения, вплоть до абсурдных. Нас же интересует коэффициент пропорциональности как показатель зависимости, не зависящий от положения начала координат. В математической статистике ликвидация такого нежелательного эффекта достигается вычитанием минимального значения, при этом самые минимальные отрицательные значения трансформируются в положительные:

$$\hat{K}_y = \frac{\bar{y} - y_{\min}}{\bar{x} - x_{\min}}; \quad \hat{K}_x = \frac{\bar{x} - x_{\min}}{\bar{y} - y_{\min}}. \quad (\text{III.2})$$

Вычисление коэффициентов пропорциональности по формуле (III.2) могло бы быть применимым и к выборкам с логарифмами значений признаков, так как если сами значения признаков будут меньше 1, то их логарифмы соответственно меньше нуля.

Однако при такой операции мы сводим любую линию зависимости к началу координат, т.е. началом координат в этом случае будет значение x_{\min} или y_{\min} или x_{\min} и y_{\min} одновременно, что явно противоречит как логике самого анализа, так и фактическим геологическим данным: практически ни одна линия зависимости при исследовании геологических объектов не проходит через начало координат.

С одной стороны, коэффициент пропорциональности нам необходим для выявления степени отклонения фактического распределения точек наблюдения от идеального пропорционального, а с другой — мы не всегда можем его вычислить.

В своих исследованиях авторы отказались от определения коэффициентов пропорциональности при определении зависимости смешанных (положительных и отрицательных) значений признаков и при определении зависимости логарифмов значений. Методы определения углов наклона линий зависимостей к координатным осям (степени отклонения фактического распределения точек наблюдения от пропорционального и, как следствие этого, — определение переменной-аргумента и переменной-функции) при логнормальной модели распределения значений признаков будут рассмотрены ниже.

III.3. Методы построения линии регрессии

Выше была рассмотрена модель пропорциональной зависимости значений изучаемых признаков. Опыт исследования реальных соотношений между значениями признаков свидетельствует о том, что строго пропорциональные взаимоотношения наблюдаются в очень редких, идеальных случаях. В то же время, изучая реальную зависимость признаков, во многих случаях полезно определять степень ее приближения к пропорциональной.

Предположим, проверкой гипотезы о "вытянутости" роя точек наблюдения на двумерной диаграмме и степени ее вытянутости (см. раздел III.1) было установлено, что линейная модель может быть использована для описания соотношений в значениях признаков. В таких ситуациях следующим этапом исследований является выбор метода построения линии регрессии, отражающей эту зависимость наилучшим образом. Следует повторить, что в настоящей работе одним из основных требований к методам построения линии регрессии является требование построения единой регрессионной прямой. Отсюда очевидно, что широко распространенные методы регрессионного анализа на основе метода наименьших квадратов здесь рассматривать нет необходимости, так как они не удовлетворяют приведенному выше требованию.

Известно, что положение вытянутого роя точек и прямой, отражающей эту вытянутость по отношению к соответствующим осям координат, определяется соотношением стандартных отклонений признаков. Обозначим соответствующие отношения через \tilde{K}_x и \tilde{K}_y ; тогда

$$\tilde{K}_x = \sigma_x / \sigma_y; \quad \tilde{K}_y = \sigma_y / \sigma_x.$$

Для демонстрации геометрического смысла приведенных отношений по аналогии с предшествующим изложением (см. раздел III.2) обозначим угол наклона продольной оси роя точек к осям координат через $\tilde{\alpha}$ — к оси абсцисс (x) и через $\tilde{\beta}$ — к оси ординат (y). Нетрудно видеть, что в этих обозначениях $\tilde{K}_y = \operatorname{tg} \tilde{\alpha}$, а $\tilde{K}_x = \operatorname{tg} \tilde{\beta}$. В соответствии с определением модели прямой пропорциональной зависимости (см. раздел III.2) продольная ось роя точек будет проходить через начало координат только в одном случае: если $\tilde{K}_x = \tilde{K}_y$ и, естественно, $\tilde{\alpha} = \tilde{\beta}$. В других обозначениях это требование можно представить следующим образом:

$$\operatorname{tg} \tilde{\alpha} = \operatorname{tg} \tilde{\beta} \quad \text{и} \quad \operatorname{tg} \tilde{\beta} = \operatorname{tg} \tilde{\alpha}.$$

Соотношение между линией пропорциональной зависимости и линией регрессии удобнее выразить через отношение угла наклона линии пропорциональной зависимости и соответствующего угла наклона роя точек. Обозначим это отношение через D_x (или D_y), тогда:

$$D_y = \operatorname{tg} \tilde{\alpha} / \operatorname{tg} \alpha \quad \text{и} \quad D_x = \operatorname{tg} \tilde{\beta} / \operatorname{tg} \beta.$$

Из приведенных выражений явно следует, что при прямой пропорциональной зависимости $D_x = D_y = 1$. При большем угле наклона оси роя точек к оси абсцисс, чем при пропорциональной зависимости $D_y > 1$, $D_x < 1$, при меньшем — $D_y < 1$, $D_x > 1$.

Возвратимся к исходным параметрам, определяющим углы наклона обсуждаемых прямых к осям координат. Получим следующие выраже-

ния для $D_x(D_y)$, которые в дальнейшем изложении будем называть "коэффициентами диспропорциональности":

$$D_x = \frac{\sigma_x/\sigma_y}{\bar{x}/\bar{y}} = \frac{\sigma_x \bar{y}}{\sigma_y \bar{x}} = \frac{\tilde{K}_x}{\hat{K}_x}$$

и

$$D_y = \frac{\sigma_y/\sigma_x}{\bar{y}/\bar{x}} = \frac{\sigma_y \bar{x}}{\sigma_x \bar{y}} = \frac{\tilde{K}_y}{\hat{K}_y}.$$

Приведенные выражения дают наглядное представление о том, что при одинаковом наклоне ряда точек к осям координат коэффициент диспропорциональности зависит от соотношения средних анализируемых признаков. И наоборот, при одинаковом соотношении средних для признаков его величина будет определяться только наклоном ряда точек к координатным осям.

Геометрический смысл $D_x(D_y)$ очевиден: степень отклонения его значений от единицы фиксирует величину отклонения зависимости значений признаков от пропорциональной. Это свойство значений коэффициента диспропорциональности нам понадобится при определении роли переменных в их взаимоотношениях (аргумент или функция) и в процедуре построения графического изображения регрессионной модели зависимости.

Из руководств по математической статистике известно, что при построении регрессионных моделей зависимости для двух случайных величин одна из них выбирается в качестве закрепленной. Закрепленная случайная величина в процедуре играет роль аргумента, в то время как другая — роль функции. В варианте регрессионного анализа, реализуемом методом наименьших квадратов, проблема выбора аргумента-функции несущественна, так как зависимость представляется с помощью построения двух линий регрессии. В этом случае каждая из случайных величин играет роль аргумента поочередно в соответствующих уравнениях регрессии и соответственно на отражающих зависимость графиках. При построении одной линии регрессии, как в нашем случае, выбор аргумента играет существенную роль. Именно в процедуре выбора аргумента с успехом используется значение коэффициента диспропорциональности.

В связи с тем что построение линии регрессии для описания зависимости между значениями признаков редко является самоцелью, способ изображения зависимости во многом определяет успешность интерпретации этой линии регрессии и, как следствие, использования ее в решении соответствующих задач. В большинстве реальных ситуаций представление зависимости признаков имеет целью анализ и сопоставление тенденций взаимного изменения значений двух признаков в разных объектах (по природе, структурному или региональному положению, возрасту и т.п.). Не менее часто возникают задачи сравнения характера поведения разных признаков в одном объекте или в объектах одного класса. Во всех подобных ситуациях естественно возникает необходимость в выявлении сопоставимости реальных зависимостей, изображенных на графике с помощью линии регрессии. Достаточно очевидно, что подобные сопо-

ставления возможны и эффективны при единообразном подходе к построению и изображению соответствующих линий регрессии.

Представляется логичным, что в качестве аргумента рационально использовать ту из анализируемых случайных величин, для которой характерна меньшая изменчивость значений. В данном случае имеется в виду не только степень рассеяния ее значений вокруг среднего, которая отображена в величине дисперсии (стандартного отклонения), но и с учетом поправки этого рассеяния на соотношение оценок средних изучаемых признаков. Другими словами, в качестве аргумента предлагается всегда использовать признак, коэффициент вариации которого меньше. Нетрудно видеть, что отношение коэффициентов вариации признаков представляет собой не что иное, как значения соответствующих коэффициентов диспропорциональности:

$$\frac{V_x}{V_y} = \frac{\sigma_x/\bar{x}}{\sigma_y/\bar{y}} = \frac{\sigma_x\bar{y}}{\sigma_y\bar{x}} = D_x \quad (\text{III.3})$$

и

$$\frac{V_y}{V_x} = \frac{\sigma_y/\bar{y}}{\sigma_x/\bar{x}} = \frac{\sigma_y\bar{x}}{\sigma_x\bar{y}} = D_y. \quad (\text{III.4})$$

Таким образом, выбор переменной, которую следует использовать в качестве аргумента, достаточно прост: в качестве аргумента всегда следует использовать значения того признака, для которого значение коэффициента диспропорциональности меньше единицы. Правило это следует из выражений (III.3 и III.4) и соответственно из соотношения коэффициентов диспропорциональности признаков

$$D_x = 1/D_y; D_y = 1/D_x.$$

Следовательно, если мы имеем дело с пропорциональной зависимостью, то оба коэффициента равны 1, в других же случаях один из них меньше, а другой — обязательно больше 1.

Как уже отмечалось (см. главу I), условию построения единой линии регрессии для отображения зависимости двух переменных отвечают методы построения прямой ортогональной регрессии. Было указано, что в основе построения линии ортогональной регрессии лежит метод построения эллипса рассеяния. Решение приведенного в цитируемом разделе уравнения для определения положения линии регрессии связано с существенными вычислительными трудностями. В данном случае для нас важно, что построение эллипса рассеяния определяется заданной доверительной вероятностью $(1 - q)$ и, следовательно, касательные к этому эллипсу, параллельные осям координат, являются не чем иным, как толерантными пределами признаков для той же вероятности $(1 - q)$. В связи с этим предлагается определять положение линии ортогональной регрессии не с помощью эллипса рассеяния, а с использованием толерантных пределов для каждого из изучаемых признаков, соответствующих заданной доверительной вероятности $(1 - q)$.

Содержание окислов щелочных металлов в вулканогенных породах
кислого состава колчеданного месторождения Молодежное (Южный Урал)

Номер п/п	Порода	Na ₂ O, %	K ₂ O, %	Номер п/п	Порода	Na ₂ O, %	K ₂ O, %
1	Дацит	3,35	0,08	9	Липарит	5,25	0,34
2	"	3,60	0,16	10	"	5,60	0,27
3	"	3,75	0,22	11	"	5,80	0,22
4	"	4,15	0,27	12	"	6,00	0,37
5	"	4,20	0,13	13	"	6,15	0,31
6	"	4,25	0,22	14	"	6,40	0,35
7	Липарит	4,80	0,18	15	"	6,60	0,40
8	"	5,10	0,28				

Рациональный способ определения толерантных пределов при симметричном и близком к нормальному распределению значений признаков (или их логарифмов) рассмотрен выше (см. раздел II.2).

Рассмотрим теперь процедуру непосредственного построения линии регрессии и продемонстрируем соотношение всех введенных коэффициентов и других параметров. В качестве примера возьмем выборку с нормальной и логнормальной моделями распределения значений признаков.

Исследуем зависимость содержаний Na₂O и K₂O в неизменных вулканогенных породах кислого состава в рудном поле колчеданного месторождения Молодежное на Южном Урале. Исходные данные по содержаниям этих окислов в породах даны в табл. 9.

В результате процедуры оценок параметров распределения получены следующие значения (Na₂O примем за x , а K₂O – за y): $n = 15$; $\bar{x} = 5,00$; $\sigma_x = 1,07$; $\bar{y} = 0,25$; $\sigma_y = 0,09$. Определим необходимые для построения линии для построения линии регрессии составляющие. В первую очередь установим роль переменных в их взаимоотношениях

$$D_x = \frac{\sigma_x \bar{y}}{\sigma_y \bar{x}}, D_x = \frac{1,07 \cdot 0,25}{0,09 \cdot 5,00} = 0,594,$$

$$D_y = \frac{\sigma_y \bar{x}}{\sigma_x \bar{y}}, D_y = \frac{0,09 \cdot 5,00}{1,07 \cdot 0,25} = 1,68,$$

но так как $D_y = \operatorname{tg} \alpha_y$, то $\alpha_y = 59,3^\circ$. Следовательно, как и было предположено, в данной ситуации в качестве аргумента необходимо использовать содержания Na₂O.

Определим толерантные пределы, задавшись доверительной вероятностью $(1 - q) = 0,95$. Как известно из предшествующего изложения (см. раздел II.2), толерантные пределы значений определяются из выражений

$$T_x = \bar{x} \pm t\sigma_x; T_y = \bar{y} \pm t\sigma_y.$$

По таблице (см. Приложение 1) для соответствующих n и $(1 - q)$ опре-

делим величину t , которая для нашего случая равна 2,15. Имея необходимые данные, вычислим размах значений для x :

$$T_{x_{\min}} = 5,00 - 2,15 \cdot 1,07 = 2,70;$$

$$T_{x_{\max}} = 5,00 + 2,15 \cdot 1,07 = 7,30$$

и для y :

$$T_{y_{\min}} = 0,25 - 2,15 \cdot 0,09 = 0,06;$$

$$T_{y_{\max}} = 0,25 + 2,15 \cdot 0,09 = 0,44.$$

Полученные параметры значений переменных позволяют нам выбрать масштаб изображения на двумерной диаграмме и провести линию зависимости (регрессии) x и y .

Как показывает практика исследований, произвольно выбирается шкала для переменной-функции, а затем через коэффициент пропорциональности рассчитывается шкала для переменной аргумента таким образом, чтобы точка пересечения значений \bar{x} и \bar{y} находилась на линии пропорциональности — биссектрисе угла, — проходящей через начало координат.

После выбора шкалы для переменной-функции и расчета шкалы значений для переменной-аргумента по параметрам распределения значений переменных проводится линия регрессии. Линия регрессии проходит (в случае прямой зависимости) через точку пересечения минимальных допустимых значений ($T_{x_{\min}}$ и $T_{y_{\min}}$), через точку пересечения средних оценок (\bar{x} и \bar{y}) и через точку пересечения максимальных допустимых ($T_{x_{\max}}$ и $T_{y_{\max}}$) значений переменных.

На рис. 6 показано положение этих линий на диаграмме $\text{Na}_2\text{O}-\text{K}_2\text{O}$. Как уже отмечалось, соотношение углов α_0 и α_y отражает степень соответствия реальной зависимости и пропорциональной зависимости; четко видно, что угол α_y больше угла α_0 , следовательно, $\text{tg } \alpha_y > \text{tg } \alpha_0$, а

$$D_y = \frac{\text{tg } \alpha_y}{\text{tg } \alpha_0} > 1; D_x = \frac{1}{D_y} < 1.$$

Что и было определено выше вычислительными процедурами на основе оценок параметров. На рис. 6 это соотношение можно видеть наглядно.

Следующей операцией процедуры исследования зависимости является определение степени разброса значений признаков вокруг построенной линии регрессии. Параметром, отражающим меру разброса точек наблюдения вокруг прямой регрессии, является условная дисперсия, которая в литературе по регрессионному анализу чаще называется остаточной дисперсией. Понятие условной, или остаточной, дисперсии и методы вычисления ее оценки хорошо известны из регрессионного анализа, реализуемого методом наименьших квадратов. Именно методом измерения расстояний от точек наблюдения до линии регрессии, кроме уже обсуждавшегося вопроса о количестве линий регрессии, отражающих зависимость двух переменных, и различаются метод наименьших квадратов и метод ортогональной регрессии.

В методе наименьших квадратов расстояние от точки наблюдения до линии регрессии измеряется параллельно соответствующей оси координат.

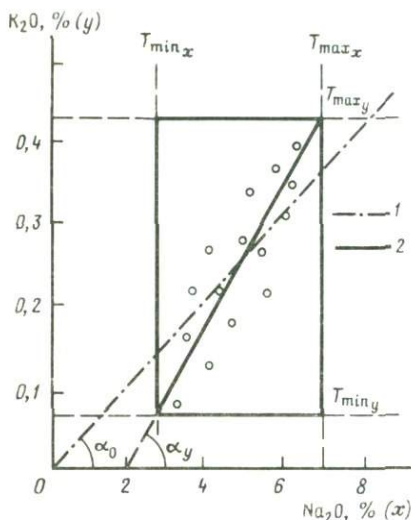


Рис. 6. Соотношение линии пропорциональности (1) и эмпирической линии регрессии (2)

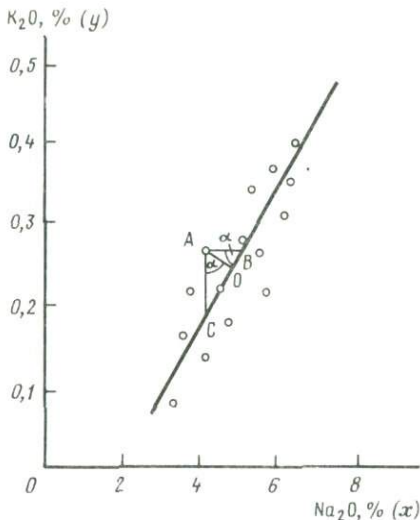


Рис. 7. Варианты определения расстояния от точки наблюдения до линии регрессии

нат. Именно эти расстояния и являются информационной базой для вычисления условной (остаточной) дисперсии. Кроме того, очевидно, что существование двух линий регрессии (x на y и y на x) определяет и необходимость вычисления двух условных (остаточных) дисперсий. При построении прямой ортогональной регрессии расстояние от точки наблюдения до прямой (одной!) измеряется по кратчайшему расстоянию (по перпендикуляру к ней, по ортогонали). Следовательно, в этом случае мы имеем лишь одно значение условной (остаточной) дисперсии.

Рассмотрим на конкретном примере соотношение указанных выше расстояний и соответственно условных (остаточных) дисперсий. Остаточная дисперсия в методе наименьших квадратов определяется с использованием выражения

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N \hat{\Delta x}_i^2, \quad (III.5)$$

где $\hat{\Delta x}_i = x_i - \hat{x}_i$; а $\hat{x}_i = \bar{x} + r_{xy} \frac{\sigma_x}{\sigma_y} (x_i - \bar{x})$, т.е. \hat{x}_i — значение x_i , полу-

ченное расчетным путем из уравнения x по y при заданном (реальном, выборочном) значении y_i . Аналогично вычисляется оценка остаточной дисперсии и для y , но определяемая относительно линии регрессии y по x . Эти расчеты производятся в случаях двух линий регрессии. В нашем случае регрессионная прямая одна. В связи с этим выражение для остаточной дисперсии имеет вид, изображенный в (III.5), лишь $\tilde{x}_i = \bar{x} + \tilde{K}_x (y_i - \bar{y})$. Аналогичное выражение используется для вычисления \tilde{y}_i .

При построении линии ортогональной регрессии расстояние от точки наблюдения до нее, т.е. $\hat{\Delta x}_i$, измеряется по прямой, перпендикулярной

линии регрессии (рис. 7). На рис. 7 четко видно, что соотношение кратчайшего расстояния от точки А до линии регрессии, т.е. $\Delta\tilde{x}_i = \Delta\tilde{y}_i$, соответствует отрезку АО. В связи с тем что угол прямой регрессии с осью абсцисс обозначен ранее через $\tilde{\alpha}$, а углы $\angle CAO$ и $\angle ABO$ также равны $\tilde{\alpha}$ (как углы с взаимно перпендикулярными сторонами), то $\Delta\tilde{x}_i = \Delta\tilde{y}_i$ (АО) можно выразить через $\Delta\tilde{x}_i$ (АВ) и $\Delta\tilde{y}_i$ (АС) следующим образом: $\Delta\tilde{x}_i = \tilde{\Delta}x_i \sin \tilde{\alpha}$; $\Delta\tilde{y}_i = \tilde{\Delta}y_i \cos \tilde{\alpha}$. Так как $\Delta\tilde{x}_i = \Delta\tilde{y}_i$, то $\tilde{\Delta}y_i \cos \tilde{\alpha} = \Delta\tilde{x}_i \sin \tilde{\alpha}$ то $\Delta\tilde{y}_i$ можно выразить следующим образом:

$$\Delta\tilde{y}_i = \Delta\tilde{x}_i \operatorname{tg} \tilde{\alpha}; \Delta\tilde{x}_i = \Delta\tilde{y}_i \operatorname{ctg} \tilde{\alpha}. \text{ С учетом того, что } \operatorname{tg} \tilde{\alpha} = \frac{\sigma_y}{\sigma_x}, \operatorname{ctg} \tilde{\alpha} = \frac{\sigma_x}{\sigma_y},$$

то $\Delta\tilde{y}_i = \Delta\tilde{x}_i \frac{\sigma_y}{\sigma_x}$; $\Delta\tilde{x}_i = \Delta\tilde{y}_i \frac{\sigma_x}{\sigma_y}$. В связи с тем, что $\Delta\tilde{x}_i = \Delta\tilde{y}_i$, то в дальнейшем из соображений удобства будем рассматривать только соотношение

$$\tilde{\Delta}y_i = \tilde{\Delta}x_i \cos \tilde{\alpha}. \quad (\text{III.6})$$

С учетом соотношения (III.6) и выражения для вычисления остаточной дисперсии (III.5) будем иметь

$$\sigma_y^2 = \frac{\cos^2 \tilde{\alpha}}{N-1} \sum_{i=1}^N \Delta\tilde{y}_i^2 \quad (\text{III.7})$$

или

$$\sigma_{\tilde{y}} = \cos \tilde{\alpha} \sqrt{\frac{1}{N-1} \sum_{i=1}^N \Delta\tilde{y}_i^2},$$

или

$$\sigma_{\tilde{y}} = \tilde{\sigma}_y \cos \tilde{\alpha}.$$

В выражении (III.7) $\sigma_{\tilde{y}}^2$ — остаточная дисперсия, вычисленная по кратчайшему расстоянию от точек наблюдения до линии регрессии.

Кратко остановимся на целях и причинах необходимости определения численного выражения остаточной дисперсии при исследовании зависимости и построения линии регрессии. В нашей методике оценка остаточной дисперсии необходима для построения зоны около линии регрессии, включающей практически всю выборочную совокупность двумерной случайной величины. Другими словами, значение оценки дисперсии предназначается для построения многоугольника, выполняющего в процедуре роль эллипса рассеяния. Вполне обоснованно можно предположить, что распределение значений расстояний точек наблюдения от линии регрессии по своим свойствам (в частности, по симметричности расположения) близко к совокупностям с нормальным распределением вероятностей. В таком случае можно использовать известное свойство дисперсии нормально распределенных случайных величин. Задав удовлетворяющее нас значение вероятности $(1 - q) = 0,95$, можно утверждать, что 95% наблюдений находится в пространстве, ограниченном прямыми, параллельными линии регрессии и расположенными от нее на расстоянии $2\sigma_{\tilde{y}}$ в каждую сторону. Естественно, что определение расстояний от каждой точки наблюдения, осуществляемое по перпендикуляру до линии регрессии, достаточно

Расчет отклонений точек наблюдения от линии регрессии
(зависимость содержаний K_2O и Na_2O в вулканогенных породах)

Номер п/п	Na_2O x_i	K_2O y_i	$x_i - \bar{x}$ Δx_i	$y_i - \bar{y}$ Δy_i	$y_i - \tilde{y}_i$ $\tilde{\Delta} y_i$
1	3,35	0,08	-1,65	-0,17	-0,03
2	3,60	0,61	-1,40	-0,09	+0,03
3	3,75	0,22	-1,25	-0,03	+0,08
4	4,15	0,27	-0,85	+0,02	+0,09
5	4,20	0,13	-0,80	-0,12	-0,05
6	4,25	0,22	-0,75	-0,03	+0,03
7	4,80	0,18	-0,20	-0,07	-0,05
8	5,10	0,28	+0,10	+0,03	+0,02
9	5,25	0,34	+0,25	+0,09	+0,07
10	5,60	0,27	+0,60	+0,02	-0,03
11	5,80	0,22	+0,80	-0,03	-0,10
12	6,00	0,37	+1,00	+0,12	+0,04
13	6,15	0,31	+1,15	+0,06	-0,04
14	6,40	0,36	+1,40	+0,10	-0,02
15	6,60	0,40	+1,60	+0,15	+0,02

трудоемко. Вычисление этих расстояний через $\tilde{\Delta}x_i$ и $\tilde{\Delta}y_i$ и тригонометрические функции ($\sin \tilde{\alpha}$ или $\cos \tilde{\alpha}$) также требует значительных затрат труда.

Рассмотрим соотношение указанных выше зон, построенных с использованием величин: $\sigma_{\tilde{x}}$, $\sigma_{\tilde{y}}$. На приведенном рис. 10 отрезки прямых имеют следующее содержание: $AO = 2\sigma_{\tilde{y}} = 2\sigma_{\tilde{x}}$; $AB = 2\tilde{\sigma}_y$; $AC = 2\tilde{\sigma}_x$. В связи с тем что $\sigma_{\tilde{x}} = \tilde{\sigma}_x \sin \tilde{\alpha}$, $\sigma_{\tilde{y}} = \tilde{\sigma}_y \cos \tilde{\alpha}$, а из рисунка очевидно, что $AO = AB \sin \tilde{\alpha} = AC \cos \tilde{\alpha}$; то прямая, проведенная через точки В, О, С, F, находится от линии регрессии в точности на требуемом расстоянии $-2\sigma_{\tilde{x}} (2\sigma_{\tilde{y}})$. Таким образом, необходимые нам прямые могут быть построены по отклонениям точек наблюдения от прямой регрессии по расстоянию как по оси абсцисс, так и по оси ординат — результат будет одинаков.

Продемонстрируем приведенное положение на реальном примере регрессии K_2O (y) по Na_2O (x). Необходимые исходные данные для расчета отклонений точек наблюдения от линии регрессии следующие:

$$\bar{x} = 5,00\%; \sigma_x = 1,07; \bar{y} = 0,25\%; \sigma_y = 0,09.$$

Расчет единичных отклонений производился по формуле

$$\tilde{\Delta}y_i = y_i - \left[\bar{y} + \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) \right]$$

или по сокращенной формуле

$$\tilde{\Delta}y_i = \Delta y_i - \frac{\sigma_y}{\sigma_x} \Delta x_i.$$

Результаты расчета отклонений точек наблюдения от линии регрессии приведены в табл. 10.

Таким образом, выполнены все расчеты для построения поля регрессии K_2O по Na_2O (y по x).

1. Параметры распределения содержаний Na_2O (x) и K_2O (y):

$$\bar{x} = 5,00\%; \quad \sigma_x = 1,07\%; \quad T_{x_{\min}} = 2,70\%; \quad T_{x_{\max}} = 7,30\%;$$

$$\bar{y} = 0,25\%; \quad \sigma_y = 0,09\%; \quad T_{y_{\min}} = 0,06\%; \quad T_{y_{\max}} = 0,44\%.$$

2. Параметры отклонений точек наблюдения от линии регрессии и величина угла наклона линии регрессии к оси абсцисс

$$\tilde{\sigma}_y = 0,055\%; \quad \tilde{T}_y = 2,15 \cdot 0,055 = 0,12\%;$$

$$D_y = 1,68; \quad \alpha_y = 59,3^\circ.$$

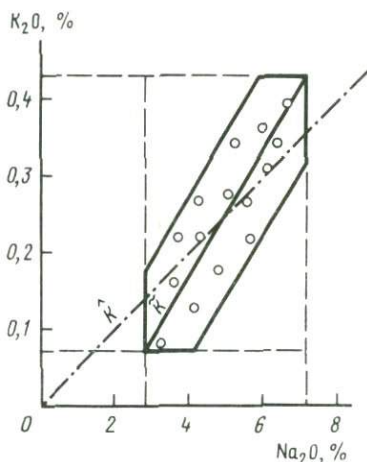


Рис. 8. Поле регрессии для содержаний K_2O и Na_2O в вулканогенных породах колчеданных месторождений

Отложив от линии регрессии по оси ординат $\pm \tilde{T}_y$, получим после регрессионной зависимости y по x , ограниченное линиями минимальных и максимальных допустимых значений y и x и линиями максимальных допустимых отклонений точек наблюдения от линии регрессии (рис. 8). Угол наклона линии регрессии соответствует коэффициенту диспропорциональности изменения значений переменных.

Параметры зависимости содержаний K_2O и Na_2O в неизменных вулканогенных породах кислого состава месторождения Молодежное (Южный Урал) даны в табл. 11. По результатам регрессионного анализа содержаний K_2O и Na_2O в вулканогенных породах можно сделать следующие выводы. Между содержаниями окислов щелочных металлов в породах наблюдается значимая положительная корреляционная зависимость. Роль переменных в их взаимоотношениях неоднозначна: Na_2O выступает в роли закрепленной переменной — аргумента, а K_2O — в роли функции. Как показывает оценка показателя эксцесса распределения, выборка относительно линии регрессии является однородной. Ниже даются уравнения регрессии для расчета содержаний K_2O по содержаниям Na_2O : $K_2O = 0,25 + 0,084(Na_2O - 5,00\%) \pm 0,11\%$; и для расчета содержаний Na_2O по содержаниям K_2O : $Na_2O = 5,00\% + 11,89(K_2O - 0,25\%) \pm 1,308\%$.

Параметры зависимости K_2O и Na_2O в вулканогенных породах кислого состава колчеданного месторождения Молодежное (Южный Урал)

Оценки параметров	Символы	Na_2O x	K_2O y
Число наблюдений	n	15	
Среднее значение	\bar{x}, \bar{y}	5,00	0,25
Стандартное отклонение от среднего значения	σ	1,07	0,09
Коэффициент пропорциональности	\hat{K}	20,00	0,05
Коэффициент регрессии	\tilde{K}	11,89	0,084
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,654	0,055
Точность линии регрессии	$\tilde{\lambda}$	$\pm 0,331$	$\pm 0,028$
Толерантные пределы отклонений от линии регрессии	T	$\pm 1,308$	$\pm 0,110$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	-0,200	
Коэффициент диспропорциональности	D	0,595	1,680
Роль переменных		Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α_y	59,3°	
Критическое значение коэффициента корреляции	$R_{0,05}(14)$	$\pm 0,506$	
Вид зависимости		Прямая	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	$\pm 0,694$	

Рассмотрим теперь определение параметров поля регрессии для переменных с логнормальной моделью распределения значений, используя для этого данные по содержаниям Pb и Zn в геохимическом ореоле Зыряновского месторождения (Рудный Алтай).

Как было определено нами ранее (см. раздел II.2), параметры распределения логарифмов содержаний свинца и цинка следующие:

$$\bar{\lg Pb} = 2,034; \sigma_{\lg Pb} = 0,493; T_{\lg Pb_{\min}} = 1,023; T_{\lg Pb_{\max}} = 3,045;$$

$$\bar{\lg Zn} = 2,511; \sigma_{\lg Zn} = 0,429; T_{\lg Zn_{\min}} = 1,627; T_{\lg Zn_{\max}} = 3,395.$$

Антилогарифм стандартного отклонения (II.2) логарифмов показывает относительный размер антилогарифмированных минимальных и максимальных допустимых значений от антилогарифмированного среднегеометрического:

$$T_{x_{\min}} = \text{antilg } \bar{\lg x} : (\text{antilg } \sigma_{\lg x})^2$$

и

$$T_{x_{\max}} = \text{antilg } \bar{\lg x} : (\text{antilg } \sigma_{\lg x})^2.$$

Расчет отклонений точек наблюдения от линии регрессии
(по оси ординат) в поле зависимости свинец–цинк

Номер п/п	lg Zn lg x	lg Pb lg y	$\Delta \lg x$	$\Delta \lg y$	$\tilde{\Delta} \lg y$
1	2	3	4	5	6
1	2,143	1,732	-0,368	-0,302	+0,121
2	1,934	1,204	-0,577	-0,830	-0,167
3	2,487	1,643	-0,024	-0,397	-0,366
4	2,378	2,276	-0,133	+0,242	+0,395
5	3,082	2,717	+0,571	+0,683	+0,027
6	2,879	2,631	+0,368	+0,597	+0,174
7	2,591	2,622	+0,080	+0,588	+0,496
8	1,792	1,732	-0,719	-0,302	+0,524
9	2,274	1,342	-0,236	-0,692	-0,421
10	1,903	1,279	-0,608	-0,755	-0,056
11	2,813	2,025	+0,302	-0,009	-0,356
12	2,813	2,079	+0,302	+0,045	-0,302
13	3,088	2,708	+0,577	+0,674	+0,011
14	2,917	2,729	+0,406	+0,695	+0,229
15	2,152	2,182	-0,359	+0,148	+0,560
16	1,954	1,820	-0,557	-0,214	+0,426
17	2,072	1,380	-0,439	-0,654	-0,150
18	2,248	2,004	-0,263	-0,030	+0,272
19	2,873	2,288	+0,362	+0,254	-0,162
20	3,160	2,480	+0,649	+0,446	-0,290
21	3,132	2,712	+0,621	+0,678	-0,036
22	2,875	2,613	+0,364	+0,579	+0,161
23	2,650	2,201	+0,139	+0,167	+0,007
24	1,954	1,431	-0,557	-0,603	+0,037
25	2,378	1,756	-0,133	-0,278	-0,125
26	2,193	1,519	-0,318	-0,515	-0,150
27	2,678	1,881	+0,176	-0,153	-0,355
28	2,886	1,959	+0,375	-0,075	-0,506

Соответственно:

$$T_{y_{\min}} = \text{antilg } \bar{\lg} y : (\text{antilg } \sigma_{\lg y})^2$$

и

$$T_{y_{\max}} = \text{antilg } \bar{\lg} y (\text{antilg } \sigma_{\lg y})^2.$$

А так как коэффициент диспропорциональности изменения значений переменных есть величина, показывающая относительный размах этих значений, то этот же D_y можно выразить

$$D_y = \frac{(\text{antilg } \sigma_{\lg y})^2}{(\text{antilg } \sigma_{\lg x})^2}$$

Параметры зависимости содержаний свинца и цинка в геохимическом ореоле
Зырянского месторождения (Рудный Алтай)

Оценка параметров	Символы	Цинк x	Свинец y
Число наблюдений	n		28
Среднее значение	$\lg x, \bar{\lg y}$	2,511	2,034
Стандартное отклонение от среднего значения	σ	0,429	0,493
Коэффициент регрессии	\tilde{K}	0,861	1,162
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,266	0,309
Точность линии регрессии	$\tilde{\lambda}$	$\pm 0,099$	$\pm 0,120$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$		+0,143
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 0,545$	$\pm 0,633$
Коэффициент диспропорциональности	D	0,75	1,343
Роль переменных		Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α_y		$53,3^\circ$
Критическое значение коэффициента корреляции	$R_{0,05}(27)$		$\pm 0,387$
Вид зависимости			Прямая
Вычисленное значение коэффициента корреляции	$R_{x,y}$		+0,709

или

$$\lg D_y = 2(\sigma_{\lg y} - \sigma_{\lg x}).$$

Таким образом, коэффициент диспропорциональности изменения переменных (тангенс угла наклона линии регрессии к оси абсцисс) может быть легко определен из параметров распределения логарифмов значений переменных (без определения коэффициента пропорциональности).

Для рассматриваемого примера зависимости содержаний свинца и цинка $\lg D_{Pb} = 2(0,493 - 0,429) = 0,128$. $\text{antilg } \lg D_{Pb} = \text{antilg } 0,128 = 1,343$. Из этого результата определяется сам угол наклона линии регрессии к оси абсцисс: $D_{Pb} = 1,343$; $\alpha_{Pb} = 53,3^\circ$.

Параметры отклонений точек наблюдения от линии зависимости определяются для логарифмов значений переменной-функции (табл. 12). Параметры поля регрессии для логарифмов содержаний свинца и цинка в геохимическом ореоле Зырянского месторождения представлены в табл. 13.

Построение поля регрессии для логарифмов значений переменных в принципе не отличается от такового для натуральных значений. Можно отметить только некоторые особенности такого построения:

1. Начало координат для переменной-функции выбирается произвольно,

но так, чтобы точка отсчета была бы меньше и минимально допустимого, и минимального реального значений.

2. Через начало координат проводится линия пропорциональности (под углом 45°).

3. Проводится линия регрессии под вычисленным углом, причем точка пересечения линий регрессии и пропорциональности является и точкой пересечения средних оценок логарифмов значений переменных.

4. На равном расстоянии от точки средних оценок проводятся линии толерантных пределов логарифмов значений (ограничительные линии пересекаются на линии регрессии) и по имеющимся минимальным и максимальным оценкам логарифмов значений переменных на координатных

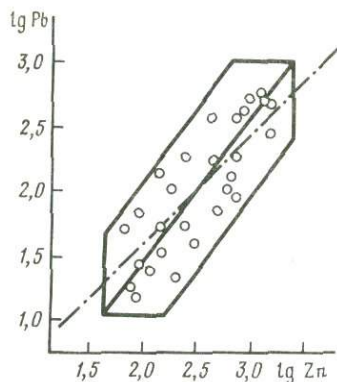


Рис. 9. Характер зависимости логарифмов содержаний Pb и Zn в геохимическом ореоле

осях рассчитываются шкалы и для переменной функции, и для переменной-аргумента.

5. По вычисленным шкалам значений и рассчитанным параметрам распределения значений и параметрам зависимости строят поле регрессии для логарифмов значений переменных.

Общий вид зависимости логарифмов содержаний свинца и цинка в геохимическом ореоле Зыряновского месторождения можно видеть из рис. 9.

III.4. Оценка корреляции поля регрессии

Необходимо остановиться еще на одном интересном и полезном следствии процедуры построения линии и поля регрессии. Известные нам параметры поля регрессии позволяют определить и численное значение коэффициента корреляции. Как будет показано ниже, эта оценка, определяемая апостериорно, заведомо несколько выше, чем вычисляемая в процедуре корреляционного анализа, но позволяет достаточно точно установить верхний предел оценки коэффициента корреляции и отсутствие ошибок в вычислениях (предварительных) значения выборочного коэффициента корреляции.

Известно [1, 4], что отношение условной дисперсии к общей дисперсии признака представляет собой квадрат корреляционного отношения. Прав-

да, в данном случае мы имеем дело с вариантом корреляционного отношения с неудовлетворяющими нас свойствами: он равен нулю при функциональной зависимости признаков и единице при отсутствии какой-либо зависимости между их значениями. Обозначим этот вариант η^* . Известен переход от этого корреляционного отношения к другому, с привычными свойствами: равенство единице при функциональной зависимости и равенстве нулю при отсутствии какой-либо зависимости. Обозначим его η , тогда

$$\eta_{x/y} = 1 - \eta_{x/y}^{*2}$$

и

$$\eta_{y/x} = 1 - \eta_{y/x}^{*2}.$$

Известно также, что при неискаженной прямолинейной зависимости между признаками значения корреляционных отношений x по y и y по x равны между собой и равные коэффициенты корреляции. Таким образом, обозначив вычисленную таким способом оценку коэффициента корреляции через $R_{x,y}$, можно записать

$$R_{x,y}^2 = 1 - \frac{\tilde{\sigma}_y^2}{\sigma_y^2}$$

или

$$R_{x,y}^2 = 1 - \frac{\tilde{\sigma}_x^2}{\sigma_x^2}.$$

В нашем примере (величины стандартных отклонений σ_x , $\tilde{\sigma}_x$, σ_y и $\tilde{\sigma}_y$ приведены выше):

$$R_{x,y}^2 \approx 1 - \frac{0,00303}{0,0081} \approx 1 - 0,373 \approx 0,626; R_{x,y} \approx 0,792;$$

$$R_{x,y}^2 \approx 1 - \frac{0,4277}{1,1445} \approx 1 - 0,373 \approx 0,626; R_{x,y} \approx 0,792.$$

Оценка коэффициента корреляции, вычисленная методом "дробового выстрела" (см. раздел III.1) равна 0,570. Естественно, что $R_{x,y}$, который является по сути своей корреляционным отношением, выше по величине, чем $R_{x,y}$ "дробового выстрела".

Следует помнить, что в данном случае оценку $R_{x,y}$ мы получаем извлечением квадратного корня из $R_{x,y}^2$, поэтому знак при коэффициенте корреляции установить невозможно. Однако он известен из предварительной проверки применимости линейной модели для описания реальной зависимости значений признаков (см. раздел III.1.).

Определение коэффициента корреляции между значениями признаков по параметрам поля регрессии можно проводить и другим способом, причем с результатом, более близким к методу "дробового выстрела", чем вышеописанный. Обратимся к рис. 9. Нетрудно понять, что чем ближе к линии регрессии расположены точки наблюдения, т. е. более вытянутый рой точек, тем сильнее сила зависимости, тем выше оценка коэффициента корреляции. Таким образом, силу зависимости между переменными можно выразить отношением ширины поля регрессии к его длине. Чем меньше это отноше-

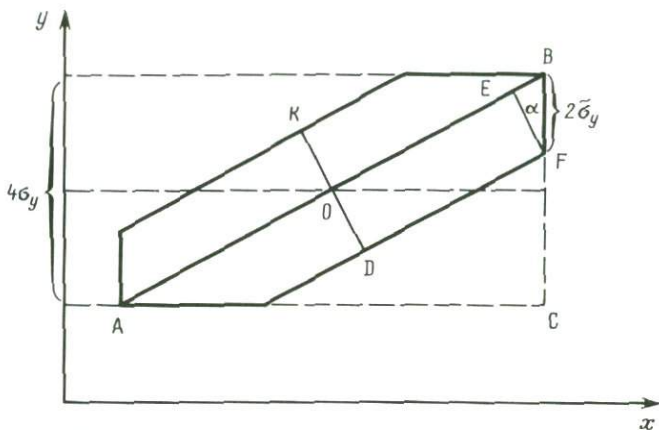


Рис. 10. Оценивание коэффициента корреляции на основе поля регрессии

ние, тем большим должна быть оценка коэффициента корреляции:

$$1 - \frac{\text{ширина поля регрессии}}{\text{длина поля регрессии}}$$

Как видно из рис. 10, ширина поля регрессии $KD = 2OD = 2EF$. Из треугольника BEF определим EF : $EF = BF \cos \alpha$, отсюда $KD = 2BF \cos \alpha$, но $BF = 2\tilde{\sigma}_y$, следовательно, $KD = 4\tilde{\sigma}_y \cos \alpha$. Длину поля регрессии AB можно

определить из треугольника ABC : $AB = \frac{BC}{\sin \alpha}$, но $BC = 4\sigma_y$, следовательно:

$AB = \frac{4\sigma_y}{\sin \alpha}$. Теперь, зная ширину и длину поля регрессии, можно определить

величину коэффициента корреляции между значениями переменных (правильнее говоря – величину коэффициента корреляции поля регрессии):

$$R_{x,y} = 1 - \frac{4\tilde{\sigma}_y \cos \alpha}{\frac{4\sigma_y}{\sin \alpha}} = 1 - \frac{\tilde{\sigma}_y}{\sigma_y} \sin \alpha \cos \alpha.$$

Так как нам известна величина $\operatorname{tg} \alpha$, то, выразив $\sin \alpha \cdot \cos \alpha$ через $\operatorname{tg} \alpha$:

$$\sin \alpha = \frac{\operatorname{tg} \alpha}{\sqrt{1 + \operatorname{tg}^2 \alpha}}; \quad \cos \alpha = \frac{1}{\sqrt{1 + \operatorname{tg}^2 \alpha}}; \quad \sin \alpha \cos \alpha = \frac{\operatorname{tg} \alpha}{1 + \operatorname{tg}^2 \alpha},$$

получим следующее выражение коэффициента корреляции:

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_y}{\sigma_y} \frac{\operatorname{tg} \alpha}{1 + \operatorname{tg}^2 \alpha}.$$

Но значение $\operatorname{tg} \alpha$ равно коэффициенту диспропорциональности изменения переменной $y - D_y$; заменив $\operatorname{tg} \alpha$ на D_y , получим окончательное выражение коэффициента корреляции значений переменных через параметры поля регрессии:

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_y}{\sigma_y} \frac{D_y}{1 + D_y^2}$$

или через параметры x :

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_x}{\sigma_x} \frac{D_x}{1 + D_x^2} .$$

При зависимости, близкой к пропорциональной, когда $D_y = 1,00$, формула вычисления коэффициента корреляции приобретает простое выражение:

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_y}{2\sigma_y} = 1 - \frac{\tilde{\sigma}_x}{2\sigma_x} .$$

Таким образом, вычисление коэффициента корреляции между значениями двух переменных не требует дополнительных расчетов, так как величина этого коэффициента определяется из параметров поля регрессии.

В своих исследованиях оценки параметров зависимости (параметров распределения точек наблюдения относительно линии регрессии), авторы определяют только для одной переменной (обычно для переменной-функции — y), причем вычисление отклонений точек наблюдения от линии регрессии производится по расстояниям по оси ординат, что значительно облегчает вычислительные операции. Оценки таковых же параметров (относительно линии регрессии) для значений переменной-аргумента (x) выводятся через коэффициент регрессии из соответствующих оценок для переменной-функции.

Процедура определения оценок параметров зависимости значений двух переменных заключается в следующих операциях.

1. Предварительное условное приращение переменным роли аргумента (x) и функции (y).
2. Оценка среднего значения каждой переменной (см. раздел III.2.).
3. Оценка стандартного отклонения значений для каждой переменной (см. раздел III.2.).
4. Оценка толерантных пределов значений переменных (см. раздел III.2.).
5. Установление пригодности линейной модели зависимости между значениями переменных (см. главу I).
6. Оценка коэффициента пропорциональности (при нормальной модели распределения значений переменных и при значениях переменных больше нуля):

$$\hat{K}_y = \frac{\bar{y}}{\bar{x}}; \quad \hat{K}_x = \frac{\bar{x}}{\bar{y}}.$$

7. Оценка коэффициента регрессии:

$$\hat{K}_y = \frac{\sigma_y}{\sigma_x}; \quad \hat{K}_x = \frac{\sigma_x}{\sigma_y}.$$

8. Оценка коэффициента диспропорциональности изменения значений переменных:

а) при нормальной модели распределения значений переменных и при значениях переменных больше нуля

$$D_y = \frac{\tilde{K}_y}{\hat{K}_y}; \quad D_x = \frac{\tilde{K}_x}{\hat{K}_x}.$$

б) при логнормальной модели распределения значений переменных

$$D_y = \text{antilg } \lg D_y; \quad \lg D_y = 2(\sigma_{\lg y} - \sigma_{\lg x}).$$

9. Окончательное определение роли переменных в их взаимоотношениях. Переменной с большим значением D придается роль функции (y), а с меньшим значением D – роль аргумента (x). При необходимости (при несовпадении предварительных и вычисленных ролей переменных) корректируются символы x и y .

10. Определение угла наклона линии регрессии к оси абсцисс:

$$\text{tg } \alpha_y = D_y; \quad \alpha_y = \text{arctg } D_y.$$

11. Определение отклонений точек наблюдения (по оси y) от линии регрессии:

$$\tilde{\Delta}y_i = \Delta y_i \mp \hat{K}_y \Delta x_i,$$

где

$$\Delta y_i = y_i - \bar{y}; \quad \Delta x_i = x_i - \bar{x}.$$

12. Оценка стандартного отклонения точек наблюдения от линии регрессии:

$$\tilde{\sigma}_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \tilde{\Delta}^2 y_i}; \quad \tilde{\sigma}_x = \tilde{K}_x \tilde{\sigma}_y.$$

13. Оценка доверительного интервала для линии регрессии:

$$\tilde{\lambda}_y = \frac{t \tilde{\sigma}_y}{\sqrt{n}}; \quad \tilde{\lambda}_x = \frac{t \tilde{\sigma}_x}{\sqrt{n}}.$$

14. Оценка толерантных пределов отклонений точек наблюдения от линии регрессии при $(1 - q)$ заданной вероятности:

$$\tilde{T}_y = \mp t \tilde{\sigma}_y; \quad \tilde{T}_x = \pm t \tilde{\sigma}_x.$$

15. Оценка показателя эксцесса распределения точек наблюдения относительно линии регрессии:

$$\tilde{\Pi}_{\text{экс}} = \frac{n(<0,5\tilde{\sigma}_j)}{0,5N} - 1,$$

где $n_{(<0,5\tilde{\sigma}_j)}$ – число наблюдений с $\tilde{\Delta}y < 0,5\tilde{\sigma}_y$.

16. Оценка коэффициента корреляции поля регрессии:

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_y}{\sigma_y} \frac{1}{1 + D_y^2}; \quad R_{x,y} = 1 - \frac{\tilde{\sigma}_x}{\sigma_x} \frac{1}{1 + D_x^2}.$$

И как итог исследования зависимости приводятся уравнения регрессии для каждой переменной, отражающие не только средние значения одной переменной при заданных значениях другой, но и допустимые отклонения вычисленных значений от средней оценки:

а) уравнение регрессии для переменной-функции:

$$\tilde{y}'_i = \bar{y} \pm \tilde{K}_y (x_i - \bar{x}) \pm \tilde{T}_y;$$

б) уравнение регрессии для переменной-аргумента:

$$\tilde{x}'_i = \bar{x} \pm \tilde{K}_x (y_i - \bar{y}) \pm \tilde{T}_x.$$

Как и в случае с оценками параметров распределения значений переменных, запись оценок параметров зависимости значений желательно производить стандартным способом. Для этой цели в Приложении 5 дан макет бланка для стандартной записи параметров зависимости значений переменных.

Глава IV

РЕГРЕССИОННЫЙ АНАЛИЗ ПРИ ИССЛЕДОВАНИИ ПОВЕДЕНИЯ ДВУХ ПРИЗНАКОВ В НЕСКОЛЬКИХ ОБЪЕКТАХ

В геологической науке и практике нередко возникают ситуации, в которых решение проблемы может быть существенно облегчено сопоставлением характера взаимного изменения значений двух признаков в изучаемом наборе объектов. Естественно, что предполагается информативность именно этой пары признаков в данной проблеме. Столь же очевидно, что исследование совокупности объектов изучения предполагает возможную их неоднородность в аспекте исследуемой проблемы, которая, по представлениям специалиста (специалистов), фиксируется характером взаимного поведения нескольких признаков, в число которых входят и два выбранных.

Как известно, геологические объекты практически во всех исследованиях характеризуются значительным числом взаимосвязанных признаков. Следует оговорить, что в данном случае имеются в виду признаки, описание которых осуществляется с помощью количественных измерений. К этой категории относятся такие признаки, как содержания петрогенных и редких химических элементов, расстояния, запасы, стоимость, координаты, разного рода индексы, показатели и т. п. Кроме того, априорно предполагается, что взаимное поведение значений пары изучаемых признаков достаточно точно описывается моделью линейной зависимости.

Для подобных ситуаций задача может быть сформулирована в общем виде следующим образом: можно ли считать, что во всех изучаемых объек-

Таблица 14

Содержания K_2O (в %) и показатель общей кислотности пород A_c
в дифференцированных базальтоидных сериях

Номер п/п	Сент-Китс		Тонга		Лассен-Парк		Бердс-Ривер		Шотландия	
	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
	1	0,65	0,17	0,81	0,38	0,77	0,72	0,68	0,51	0,77
2	0,74	0,30	0,83	0,37	0,96	1,49	0,68	0,58	0,82	1,39
3	0,80	0,40	0,86	0,20	0,99	1,52	0,79	0,63	0,87	1,67
4	0,85	0,44	0,86	0,31	1,02	1,45	0,83	1,11	0,92	2,12
5	0,91	0,46	0,88	0,46	1,04	1,57	1,03	1,79	1,03	2,36
6	1,00	0,51	0,91	0,33	1,07	1,51	1,12	2,32	1,16	2,70
7	1,08	0,51	0,91	0,38	1,25	2,03	1,32	2,73		
8	1,15	0,63	0,92	0,29	1,34	2,53				
9	1,24	0,62	0,95	0,51	1,35	2,16				
10	1,36	0,66	0,98	0,22	1,42	2,34				
11			0,99	0,58	1,43	2,41				
12			1,01	0,64	1,46	2,39				
13			1,03	0,58						
14			1,10	0,48						
15			1,17	0,76						
16			1,22	0,75						
17			1,34	0,92						
18			1,34	0,94						
19			1,38	0,94						
20			1,38	0,94						
21			1,42	0,85						
22			1,49	0,98						
23			1,67	0,96						
24			1,82	1,22						
25			1,85	1,14						

Номер п/п	Анды (Чили)		Анды (Перу)		Восточная Ирландия		Исландия		Восточно-Афри- канский Рифт	
	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
	1	0,74	1,41	0,86	2,28	0,61	1,25	0,63	0,19	0,54
2	0,76	1,25	0,87	1,91	0,64	1,29	0,63	0,22	0,60	0,95
3	0,80	0,91	0,87	1,97	0,66	1,06	0,65	0,28	0,69	1,90
4	0,82	1,33	0,91	2,32	0,66	1,15	0,66	0,22	0,76	2,51
5	0,82	1,41	0,92	1,91	0,68	2,00	0,67	0,53	0,95	4,03
6	0,89	1,49	0,93	1,87	0,72	1,68	0,68	0,38		
7	0,94	1,49	0,95	1,99	0,76	1,19	0,69	0,49		
8	0,95	1,33	0,95	2,49	0,77	1,67	0,71	0,56		
9	0,95	1,49	0,96	1,91	0,82	2,15	0,72	0,61		
10	0,95	1,66	0,99	2,41	0,83	1,88	0,78	1,01		
11	0,96	1,74	1,01	2,74	0,85	2,29	0,82	0,89		
12	0,97	2,41	1,02	2,57	0,92	3,15	0,92	1,49		
13	0,98	1,49	1,02	2,57	1,17	3,23	1,09	1,65		
14	0,98	2,32	1,05	2,74	1,21	2,36	1,22	2,02		

Таблица 14 (окончание)

Номер п/п	Анды (Чили)		Анды (Перу)		Восточная Ирландия		Исландия		Восточно-Афри- канский Рифт	
	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
15	1,01	1,91	1,08	2,49	1,22	2,05	1,53	2,81		
16	1,04	2,16	1,08	2,57	1,64	5,30	1,61	2,99		
17	1,05	1,66	1,08	2,82			1,62	2,83		
18	1,08	2,57	1,09	2,74						
19	1,10	2,16	1,11	2,74						
20	1,11	1,58	1,37	3,59						
21	1,11	2,24	1,40	3,75						
22	1,14	2,57	1,42	3,73						
23	1,18	2,08	1,43	3,93						
24	1,24	2,49								

Номер п/п	Бойна		Пасха		Кения		Джебел-Абиад		Святая Елена	
	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
1	0,61	0,66	0,64	0,61	0,62	0,54	0,56	1,08	0,56	1,33
2	0,62	0,91	0,65	0,22	0,62	0,79	0,60	0,45	0,61	0,95
3	0,64	0,52	0,67	0,26	0,62	0,98	0,61	0,88	0,64	1,50
4	0,64	0,91	0,68	0,27	0,62	1,00	0,64	1,20	0,80	2,40
5	0,64	1,12	0,68	0,58	0,63	0,59	0,67	1,37	0,85	3,34
6	0,66	1,29	0,70	0,52	0,63	0,66	0,78	2,60	0,86	2,68
7	0,88	1,78	0,71	0,32	0,63	0,95	0,79	2,08	0,89	4,12
8	0,95	2,71	0,73	0,90	0,63	0,96	0,83	3,53	0,91	3,54
9	1,00	2,21	0,82	1,19	0,64	0,89	0,84	3,00	0,91	3,90
10	1,13	3,57	0,96	1,57	0,64	1,02	0,84	3,01		
11	1,25	3,74	1,06	1,81	0,65	0,67	0,85	3,21		
12	1,29	3,75	1,08	1,82	0,85	3,35	0,86	2,85		
13	1,32	3,64	1,40	3,16	0,86	3,40	0,89	4,33		
14	1,35	3,70	1,43	3,22	0,87	3,03	0,90	3,92		
15	1,40	3,64	1,43	3,26	0,93	4,31	0,91	3,95		
16	1,41	3,65			0,94	3,92	0,92	4,35		
17	1,43	3,64			0,94	3,93	0,92	4,37		
18					0,94	4,34	0,92	4,42		
19					0,96	4,45	0,93	4,44		
20					0,97	4,28	0,95	4,16		
21					0,98	4,33	0,95	4,22		

Таблица 15

Параметры распределения содержаний K_2O (%) и значений показателя общей кислотности A_c в дифференцированных базальтоидных сериях

Статистические оценки	Символы	Сент-Китс		Тонга	
		A_c	K_2O	A_c	K_2O
		Число наблюдений	n	10	
Среднее значение	\bar{x}, \bar{y}	0,98	0,47	1,17	0,65
Стандартное отклонение	σ	0,23	0,15	0,31	0,30
Доверительный интервал для среднего	λ	$\pm 0,16$	$\pm 0,11$	$\pm 0,13$	$\pm 0,13$
Показатель асимметрии распределения	$P_{ас}$	0,00	0,00	-0,12	-0,12
Показатель эксцесса распределения	$P_{экс}$	-0,40	0,00	-0,60	-0,52
Толерантные пределы значений при $(1 - q) = 0,90$.					
а) минимальные	T_{mix}	0,46	0,12	0,52	0,02
б) максимальные	T_{max}	1,50	0,82	1,81	1,27

Статистические оценки	Символы	Восточная Ирландия		Исландия		Бойна	
		A_c	K_2O	A_c	K_2O	A_c	K_2O
		Число наблюдений	n	16		17	
Среднее значение	\bar{x}, \bar{y}	0,89	2,11	0,92	1,13	1,01	2,44
Стандартное отклонение	σ	0,28	1,08	0,36	0,99	0,33	1,34
Доверительный интервал для среднего	λ	$\pm 0,15$	$\pm 0,57$	$\pm 0,18$	$\pm 0,51$	$\pm 0,17$	$\pm 0,67$
Показатель асимметрии распределения	$P_{ас}$	-0,38	-0,25	-0,29	-0,29	-0,06	+0,06
Показатель эксцесса распределения	$P_{экс}$	-0,25	-0,00	-0,53	-0,65	-0,53	-0,77
Толерантные пределы значений при $(1 - q) = 0,90$.							
а) минимальные	T_{min}	0,28	-0,18	0,16	-0,97	0,32	-0,33
б) максимальные	T_{max}	1,49	4,40	1,68	3,22	1,71	5,20

тах градиент изменения концентраций данных признаков сохраняется неизменным, что отражает представление об однотипном поведении определяемых элементов в процессе формирования изучаемых объектов.

При реализации задач подобного содержания в качестве эффективного, наглядного и рационального (с учетом значимости получаемой информации и затрат труда на реализацию) метода предлагается применение модифицированной методики построения прямых ортогональной регрессии, отображающих характер изменения концентраций пары изучаемых признаков во всех объектах и представление их на одной двумерной диаграмме.

Лассен-Парк		Шотландия		Бердс-Ривер		Анды (Чили)		Анды (Перу)	
A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
12		6		7		24		23	
1,18	1,84	0,93	1,89	0,92	1,38	0,98	1,80	1,06	2,61
0,23	0,55	0,14	0,62	0,24	0,90	0,13	0,47	0,18	0,62
$\pm 0,14$	$\pm 0,35$	$\pm 0,15$	$\pm 0,65$	$\pm 0,22$	$\pm 0,84$	$\pm 0,06$	$\pm 0,20$	$\pm 0,08$	$\pm 0,27$
0,00	0,00	-0,33	0,00	-0,14	-0,14	-0,17	-0,17	-0,22	-0,22
-0,67	-0,67	-0,33	-0,33	-0,43	-0,43	-0,08	-0,50	-0,13	+0,04
0,68	0,64	0,56	0,30	0,33	-0,83	0,71	0,82	0,69	1,33
1,67	3,05	1,30	3,47	1,51	3,60	1,26	2,78	1,43	3,89

Пасха		Кения		Джебел-Абиад		Святая Елена		Восточно-Африканский Рифт	
A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
15		21		21		9		5	
0,91	1,31	0,77	2,30	0,82	3,02	0,78	2,64	0,71	2,01
0,30	1,12	0,15	1,63	0,13	1,34	0,14	1,17	0,16	1,35
$\pm 0,17$	$\pm 0,62$	$\pm 0,07$	$\pm 0,75$	$\pm 0,06$	$\pm 0,61$	$\pm 0,11$	$\pm 0,90$	$\pm 0,20$	$\pm 1,68$
-0,20	-0,20	-0,05	-0,05	+0,33	+0,05	+0,33	+0,11	-0,20	-0,20
-0,73	-0,33	-1,00	-0,95	-0,33	-0,43	-0,56	-0,56	-0,20	-0,20
0,26	-1,10	0,45	-1,11	0,55	0,21	0,46	-0,07	-0,27	-1,74
1,55	3,73	1,09	5,72	1,08	5,83	1,10	5,35	1,15	5,76

Некоторые необходимые пояснения. Естественно, что в данном случае для всех объектов в качестве аргумента используются значения одного и того же признака. Кроме того, масштаб изображения зависимости должен быть приведен к единому числовому выражению. Процедуру реализации подобной задачи имеет смысл рассмотреть на конкретном примере.

В качестве примера проведения регрессионного анализа двух переменных (признаков) с целью сравнения градиентов их изменения в совокупности геологических объектов рассмотрим исследование зависимости содержаний K_2O от значений показателя кислотности магм A_c [6]. Выборочные

Таблица 16

Оценки параметров зависимости содержаний K_2O (в %) и значений показателя общей кислотности A_c в дифференцированных базальтоидных сериях

Статистические оценки	Символы	Сент-Китс		Тонга	
		A_c	K_2O	A_c	K_2O
Число наблюдений	n	10		25	
Среднее значение	\bar{x}, \bar{y}	0,98	0,47	1,17	0,65
Стандартное отклонение	σ	0,23	0,15	0,31	0,30
Коэффициент пропорциональности	\hat{K}	2,08	0,48	1,81	0,55
Коэффициент регрессии	\tilde{K}	1,49	0,67	1,03	0,97
Стандартное отклонение точек наблюдения от линии регрессии	$\tilde{\sigma}$	0,07	0,05	0,11	0,11
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,05$	$\pm 0,03$	$\pm 0,05$	$\pm 0,04$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{экс}$	+0,80		+0,52	
Толерантные пределы отклонений точек наблюдения от линии регрессии	\tilde{T}	$\pm 0,14$	$\pm 0,10$	$\pm 0,21$	$\pm 0,22$
Коэффициент диспропорциональности	D	0,72	1,40	0,57	1,76
Роль переменных		Аргумент Функция		Аргумент Функция	
Угол наклона линии регрессии к оси абсцисс	α	54,4°		60,4°	
Коэффициент корреляции поля регрессии	R	+0,840		+0,820	
Критическое значение коэффициента корреляции	$R_{крит}$	0,653		0,400	
Зависимость		Прямая		Прямая	

совокупности, участвующие в процедуре, представляют дифференцированные базальтоидные серии.

В основу исследования положено обобщение современных данных по дифференцированным базальтоидным сериям, представляющим различные в тектоническом отношении петрографические провинции. Известно, что по поведению K_2O в процессе дифференциации исходной магмы можно различать, по крайней мере, два основных петрологических типа базальтовых серий: толеитовые и щелочно-базальтовые серии. В связи с этим градиент изменения содержаний K_2O в процессе дифференциации магмы во многом определяет принадлежность изучаемого объекта к одному из основных типов базальтовых серий.

Для сравнения характера изменения содержаний K_2O в качестве единого показателя дифференциации вещества использовался показатель общей кислотности магм A_c , предложенный Л.С. Бородиным [6]. Показатель A_c своими значениями отражает динамику изменения соотношения кремнезема и остальных петрогенных химических элементов в процессе дифферен-

Лассен-Парк		Бердс-Ривер		Шотландия		Анды (Чили)	
A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
	12		7		6		24
1,18	1,84	0,92	1,38	0,93	1,89	0,98	1,80
0,23	0,55	0,24	0,90	0,14	0,62	0,13	0,46
0,64	1,57	0,67	1,50	0,49	2,03	0,55	1,83
0,41	2,41	0,27	3,73	0,23	4,29	0,28	3,58
0,06	0,14	0,04	0,15	0,04	0,16	0,09	0,33
$\pm 0,03$	$\pm 0,08$	$\pm 0,03$	$\pm 0,11$	$\pm 0,04$	$\pm 0,17$	$\pm 0,04$	$\pm 0,13$
	+0,33		+0,14		+0,00		+0,25
$\pm 0,11$	$\pm 0,26$	$\pm 0,08$	$\pm 0,30$	$\pm 0,10$	$\pm 0,41$	$\pm 0,18$	$\pm 0,64$
0,65	1,54	0,40	2,49	0,47	2,11	0,51	1,96
Аргумент	Функция	Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
	$57,0^\circ$		$68,1^\circ$		$64,7^\circ$		$62,9^\circ$
	+0,880		+0,920		+0,870		+0,550
	0,590		0,800		0,820		0,410
	Прямая		Прямая		Прямая		Прямая

циации магмы. С использованием материалов совместно с Л.С. Бородиным [7] изучалось изменение содержаний K_2O в связи с изменением показателя кислотности магм A_c и базальтоидные серии сравнивались по градиенту изменения содержаний K_2O в процессе магматической глубинной дифференциации.

Исследуемые выборочные совокупности представляют следующие базальтоидные серии: Сент-Китс, Тонга, Лассен-Парк, Восточная Ирландия, Исландия, Анды перуанские, Анды чилийские, Шотландия, Бердс-Ривер, Бойна, Святая Елена, Пасха, Кения, Джебел-Абиад, Восточно-Африканский Рифт. Исходные данные по содержаниям окиси калия и по значениям A_c перечисленных выше серий приведены в табл. 14.

Регрессионный анализ зависимости K_2O-A_c осуществлялся поэтапно. На первом этапе вычислялись оценки параметров распределения содержаний K_2O и значений A_c в частных выборках (сериях), результаты которого представлены в табл. 15. На втором этапе вычислялись оценки параметров зависимости содержаний K_2O и значений A_c (табл. 16).

Таблица 16 (продолжение)

Статистические оценки	Символы	Анды (Перу)		Восточная Ирландия	
		A_c	K_2O	A_c	K_2O
Число наблюдений	n	23		16	
Среднее значение	\bar{x}, \bar{y}	1,06	2,61	0,89	2,11
Стандартное отклонение	σ	0,18	0,62	0,29	1,08
Коэффициент пропорциональности	\tilde{K}	0,41	2,46	0,42	2,38
Коэффициент регрессии	\tilde{K}	0,29	3,48	0,27	3,78
Стандартное отклонение точек наблюдения от линии регрессии	$\tilde{\sigma}$	0,06	0,19	0,15	0,55
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,02$	$\pm 0,08$	$\pm 0,07$	$\pm 0,27$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	+0,39		0,00	
Толерантные пределы отклонений точек наблюдения от линии регрессии	\tilde{T}	$\pm 0,11$	$\pm 0,38$	$\pm 0,29$	$\pm 1,09$
Коэффициент диспропорциональности	D	0,71	1,41	0,63	1,59
Роль переменных		Аргумент Функция		Аргумент Функция	
Угол наклона линии регрессии к оси абсцисс	α	54,7°		57,8°	
Коэффициент корреляции поля регрессии	R	+0,850		+0,740	
Критическое значение коэффициента корреляции	$R_{\text{крит}}$	0,420		0,510	
Зависимость		Прямая		Прямая	

Прежде чем перейти к непосредственному рассмотрению результатов применения методики для описания зависимости содержания K_2O от значений кислотности магмы A_c , следует остановиться на некоторых интересных и важных аспектах как самой методики, так и ее применения. При обсуждении методики регрессионного и корреляционного анализа указывалось (см. главу III), что обязательным для выполнения требований математической статистики к анализу зависимости значений признаков является требование однородности двумерной выборочной совокупности. Как известно, одним из показателей однородности выборки является значение коэффициента вариации. В табл. 17 приведены значения эмпирического коэффициента вариации для переменных x (A_c) и y (K_2O) в базальтоидных сериях. Выше было приведено допустимое значение коэффициента вариации для однородных совокупностей с симметричным распределением (0,60), превышение которого свидетельствует о существенной неоднородности изучаемой совокупности. Нетрудно видеть из приве-

Исландия		Бойна		Пасха		Кения	
A_c	K_2O	A_c	K_2O	A_c	K_2O	A_c	K_2O
17		17		15		21	
0,92	1,13	1,01	2,44	0,91	1,31	0,77	2,30
0,36	0,99	0,33	1,30	0,30	1,12	0,15	1,63
0,82	1,30	0,42	2,41	0,69	1,45	0,33	2,99
0,36	2,76	0,25	3,98	0,27	3,74	0,09	10,61
0,05	0,15	0,08	0,33	0,04	0,17	0,02	0,21
$\pm 0,03$	$\pm 0,07$	$\pm 0,04$	$\pm 0,16$	$\pm 0,02$	$\pm 0,08$	$\pm 0,01$	$\pm 0,09$
+0,06		+0,41		+0,20		+0,74	
$\pm 0,10$	$\pm 0,29$	$\pm 0,16$	$\pm 0,64$	$\pm 0,09$	$\pm 0,32$	$\pm 0,04$	$\pm 0,42$
0,44	2,25	0,60	1,67	0,39	2,59	0,28	3,55
Аргумент	Функция	Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
66,1°		58,9°		68,9°		74,3°	
+0,930		+0,870		+0,930		+0,940	
0,490		0,490		0,520		+0,940	
Прямая		Прямая		Прямая		Прямая	

денной таблицы, что в одном случае для переменной x и в пяти случаях — для y значение коэффициента вариации превышает допустимое. Следовательно, из 15 выборок мы должны исключить из рассмотрения шесть, как неудовлетворяющие требование однородности. Здесь необходимо отступление от основного изложения, которое позволит более наглядно показать особенности предлагаемого методики.

В математической статистике регрессионный анализ как на основе метода наименьших квадратов, так и с помощью ортогональной регрессии в качестве основного параметра использует коэффициент корреляции, который обычно определяется как "парный коэффициент корреляции". Выражение для вычисления и смыслов содержание этого статистического параметра в условиях выполнения требования однородности анализируемой совокупности можно найти в любом руководстве по математической статистике. Известны [5, 10] свойства оценки парного коэффициента корреляции и в условиях неоднородных выборочных совокупностей. В

Таблица 16 (окончание)

Статистические оценки	Символы	Джебел-Абиад		Святая Елена		Восточно-Африканский Рифт	
		A_c	$K_2 O$	A_c	$K_2 O$	A_c	$K_2 O$
Число наблюдений	n	21		9		5	
Среднее значение	\bar{x}, \bar{y}	0,82	3,02	0,78	2,64	0,71	2,01
Стандартное отклонение	σ	0,13	1,34	0,14	1,17	0,16	1,35
Коэффициент пропорциональности	\hat{K}	0,27	3,70	0,30	3,38	0,35	2,84
Коэффициент регрессии	\tilde{K}	0,09	10,68	0,12	8,44	0,12	8,48
Стандартное отклонение точек наблюдения от линии регрессии	$\tilde{\sigma}$	0,03	0,36	0,05	0,41	0,01	0,09
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,01$	$\pm 0,15$	$\pm 0,03$	$\pm 0,27$	$\pm 0,01$	$\pm 0,08$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	+0,33		+0,11		+0,20	
Толерантные пределы отклонений точек наблюдения от линии регрессии	\tilde{T}	$\pm 0,07$	$\pm 0,70$	$\pm 0,09$	$\pm 0,80$	$\pm 0,02$	$\pm 0,18$
Коэффициент диспропорциональности	D	0,35	2,89	0,40	2,50	0,33	2,99
Роль переменных		Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	70,9°		68,2°		71,5°	
Коэффициент корреляции поля регрессии	R	+0,870		+0,830		+0,970	
Критическое значение коэффициента корреляции	$R_{\text{крит}}$	0,440		0,690		0,960	
Зависимость		Прямая		Прямая		Прямая	

кратком изложении эти свойства можно определить следующим образом.

Если анализируемая выборочная совокупность состоит из двух подсовокупностей с разными значениями двумерных средних, то значение оценки парного коэффициента корреляции отражает не степень компактности точек наблюдения вокруг некоторой прямой (как в классическом случае однородных совокупностей), а величину расстояния между средними

Значения коэффициента вариации для переменных x (A_c) и y (K_2O), вычисленные для выборок базальтоидных серий (в долях единицы)

Номер п/п	Базальтоидная серия	Коэффициент вариации	
		по x	по y
1	Сент-Китс	0,23	0,32
2	Тонга	0,26	0,46
3	Лассен-Парк	0,19	0,30
4	Шотландия	1,22	0,33
5	Бердс-Ривер	0,26	0,65
6	Анды чилийские	0,13	0,26
7	Анды перуанские	0,17	0,24
8	Восточная Ирландия	0,31	0,51
9	Исландия	0,39	0,88
10	Бойна	0,32	0,55
11	Пасха	0,33	0,85
12	Кения	0,19	0,71
13	Джебел-Абиад	0,16	0,44
14	Святая Елена	0,18	0,44
15	Восточно-Африканский Рифт	0,22	0,67

подсовокупностей даже в тех случаях, когда в каждой подсовокупности отсутствует какая-либо зависимость между признаками. Наличие более чем двух подсовокупностей, объединенных в исследуемую выборку, влияет на значение оценки парного коэффициента корреляции аналогично ситуации с выборками малого объема (3–5 наблюдений). Величина оценки зависит от взаимного расположения двумерных средних подсовокупностей: она будет тем больше, чем ближе к прямой линии располагаются точки средних на двумерной диаграмме.

Необходимо остановиться на еще одном вопросе, связанном со свойствами парного коэффициента корреляции. Особенно сложной является ситуация, когда две подсовокупности неоднородной выборки имеют резко различные объемы наблюдений. В подобном случае из-за двух-трех наблюдений, резко отклоняющихся по значениям признаков от основной совокупности, исследователь получит искаженное, не соответствующее форме роя основной части выборки значение оценки коэффициента корреляции. При применении рассматриваемой методики нас не может устроить ни ситуация, когда оценка коэффициента корреляции близка к нулю при вытянутой в одном направлении форме роя и двух-трех точках, удаленных в стороны от него, ни ситуация, когда основная совокупность образует круг на двумерной диаграмме, т.е. в ней отсутствует какая-либо зависимость, а две-три точки, удаленные от него в одном направлении, определяют вычисленную величину оценки коэффициента корреляции, близкую к единице. Особенно неудобен с позиций рассматриваемой методики второй случай. Естественно, что строить линию регрессии любым из известных

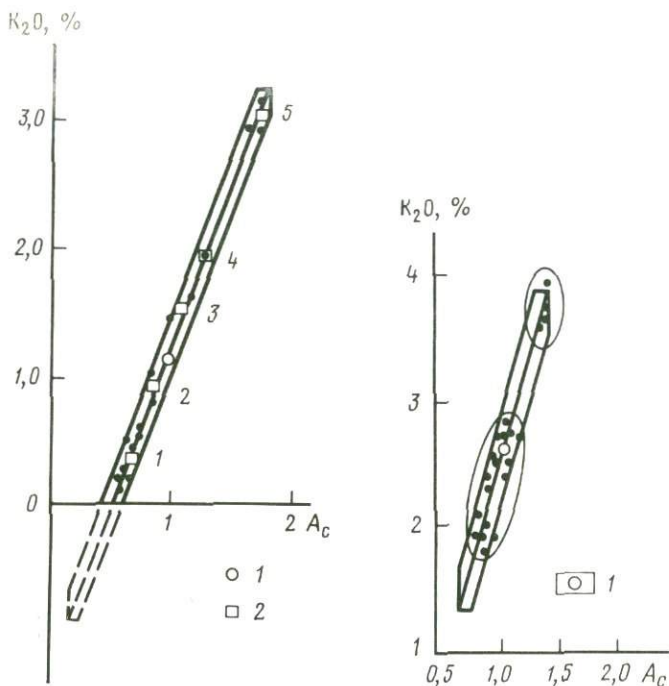


Рис. 11. Подразделение выборки серии пород Исландия на подсовокупности
1 – среднее по выборке; 2 – средние по подсовокупностям

Рис. 12. Подсовокупности в выборке пород Анды (Перу)
1 – двумерное среднее $\{\bar{x}, \bar{y}\}$ по выборке

способов нет оснований. Как поступить в такой сложной ситуации: значение коэффициента корреляции близко к единице, а построение линии регрессии безосновательно? В такой ситуации максимально эффективным выходом является применение рассмотренного выше метода вычисления коэффициента корреляции "дробового выстрела". Его оценка, в отличие от оценки парного коэффициента корреляции в этом случае будет близка к нулю, т.е. исследователь будет знать априори, что по имеющимся в его распоряжении данным нет оснований для построения линии регрессии.

В ситуациях, когда объемы подсовокупностей примерно равны или сопоставимы, то оба указанных выше коэффициента корреляции приводят к близкому результату – величина их оценок определяется одним и тем же фактором: положением средних для подсовокупностей на двумерной диаграмме.

Рассмотрим приведенные выше особенности эмпирических значений коэффициентов корреляции на примере выборки базальтоидной серии Исландии, для которой (см. табл. 17) вычисленным значением коэффициента вариации установлена неоднородность выборочной совокупности по переменной y ($V_y = 0,88 > 0,6$). Точечная диаграмма наблюдений в координатах A_c (x – аргумент)– K_2O (y – функция) приведена на рис. 11.

Подсовокупности общей выборки базальтоидной серии Исландия

Номер п/п	A_c x_i	$K_2 O$ y_i	\bar{x}	\bar{y}
1-я подсовокупность, $n_1 = 9$				
1	0,63	0,19		
2	0,66	0,22		
3	0,67	0,53		
4	0,65	0,28		
5	0,63	0,22		
6	0,68	0,38		
7	0,71	0,56		
8	0,69	0,49		
9	0,72	0,61	0,67	0,39
2-я подсовокупность, $n_2 = 2$				
10	0,78	1,01		
11	0,82	0,89	0,80	0,95
3-я подсовокупность, $n_3 = 2$				
12	0,93	1,49		
13	1,09	1,65	1,01	1,57
4-я подсовокупность, $n_4 = 1$				
14	1,22	2,02	1,22	2,02
5-я подсовокупность, $n_5 = 3$				
15	1,62	2,83		
16	1,61	2,99		
17	1,53	2,81	1,59	2,88

На приведенной точечной диаграмме с достаточной очевидностью фиксируются несколько сгущений точек, обособленных друг от друга. В данном случае условно можно выделить пять подсовокупностей, которым на графике (см. рис. 11) приписаны номера от 1 до 5, а соответствующие оценки средних приведены в табл. 18. Положение линии регрессии, построенной в соответствии с рассматриваемой методикой, и средних указанных выше подсовокупностей, вынесенных над диаграмму, свидетельствует, что средние подсовокупностей не только расположены близ линии регрессии, но и подтверждают положение, что при наличии нескольких подсовокупностей линия регрессии определяется положением их средних.

Вполне очевидно, что в данном случае коэффициент парной корреляции, а равно и коэффициент корреляции "дробового выстрела" в одинаковой степени отражает ориентированное вдоль прямой положение двумерных средних подсовокупностей, составляющих общую выборку базальтоидной серии Исландия. Именно это следует из сопоставления эмпирических значений коэффициентов корреляции, вычисленных различными способами (рассматриваемыми в данной работе) для анализируемых выборок базальтоидных серий (табл. 19). Для обсуждаемой здесь выборки Исландия значения коэффициентов корреляции очень близки (0,937 и 1,000). Причем

Оценки коэффициента корреляции для значений переменных x (A_c) и y ($K_2 O$), вычисленные разными способами

Серия	Число наблюдений	Коэффициент корреляции			
		допустимый	парный	"дробового выстрела"	поля регрессии
Сент-Китс	10	0,632	0,949	1,000	0,943
Тонга	25	0,400	0,929	0,760	0,930
Лассен-Парк	12	0,576	0,970	1,000	0,967
Бердс-Ривер	7	0,754	0,986	1,000	0,986
Шотландия	6	0,811	0,967	1,000	0,966
Анды (Чили)	24	0,410	0,761	0,667	0,712
Анды (Перу)	23	0,420	0,952	0,652	0,952
Восточная Ирландия	16	0,497	0,693	0,750	0,861
Исландия	17	0,482	0,937	1,000	0,988
Восточно-Африканский Рифт	5	0,878	0,998	1,000	0,998
Бойна	17	0,482	0,968	1,000	0,967
Пасха	15	0,514	0,989	1,000	0,988
Кения	21	0,433	0,991	0,810	0,992
Джебел-Абиад	21	0,433	0,965	0,714	0,963
Святая Елена	9	0,666	0,940	1,000	0,936

значение оценки коэффициента корреляции "дробового выстрела" не изменится и при условии, если его вычислить только по значениям пяти средних для подсовокупностей. В этом нетрудно убедиться, используя график с положением двумерных средних подсовокупностей (см. рис. 11).

Из приведенной таблицы оценок коэффициентов корреляции, вычисленных разными способами (см. табл. 19), очевидно, что в большинстве случаев оценки парного коэффициента корреляции и вычисленные методом "дробового выстрела" практически сопоставимы по величине. Однако в некоторых случаях значения коэффициента корреляции "дробового выстрела" существенно ниже, чем парного (серии Тонга, Анды чилийские, Джебел-Абиад). В этих случаях мы имеем дело с ситуацией двух подсовокупностей с различным объемом наблюдений. Причины подобного соотношения оценок коэффициента корреляции и подобные ситуации вообще можно наглядно продемонстрировать на примере серии Анды перуанские.

На точечной диаграмме (см. рис. 12) отчетливо видно, что выборка состоит из двух подсовокупностей с разным числом наблюдений ($n_1 = 19$, $n_2 = 4$). Однако этот случай не относится к предельным, рассмотренным выше, так как в большой по объему подсовокупности рой точек наблюдения имеет явную вытянутость, о чем свидетельствует оценка коэффициента корреляции "дробового выстрела", вычисленная для нее $R_{x,y} = 0,789$ при допустимом значении $R_{0,05}(22) = 0,450$. В связи с этим и оценка коэффициента корреляции "дробового выстрела" по всей выборке (0,652) больше допустимого значения (0,409), что позволяет нам все же строить линию регрессии, которая, естественно, в большей степени отражает

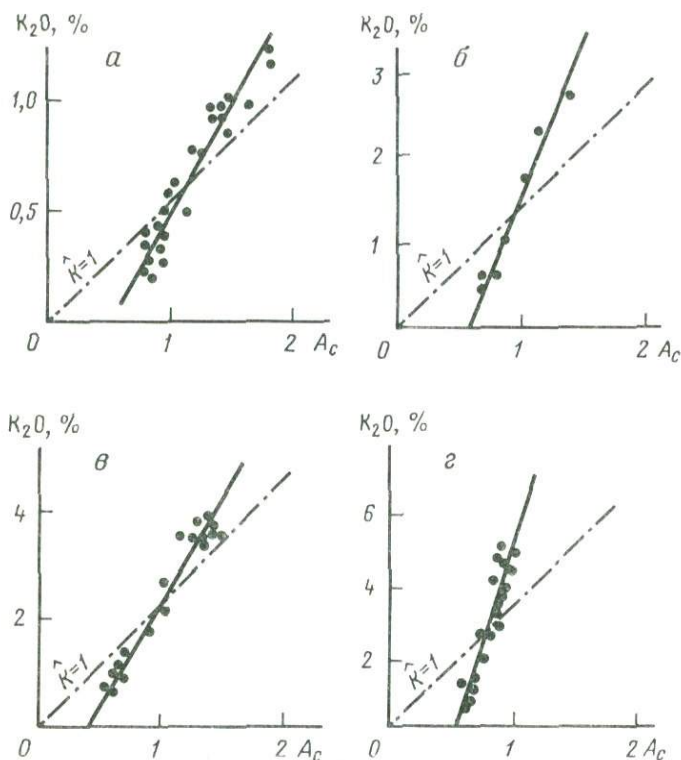


Рис. 13. Линии регрессии содержаний K_2O и A_c для дифференцированных базальтоидных серий:

a – Тонга, *б* – Бердс-Ривер, *в* – Бойна, *г* – Джебел-Абияд

зависимость между значениями признаков в подсовокупности с преобладающей в общей выборке долей наблюдений. В результате предварительного статистического анализа имеющихся данных установлено, что значения показателя однородности (см. табл. 17) и положение точек на двумерных диаграммах в координатах K_2O-A_c для ряда выборок свидетельствуют о неоднородности последних. Наблюдаемая неоднородность обусловлена либо различной степенью опробованности тех или иных разновидностей пород, либо различиями в их распространенности в пределах серий. Использование таких выборок для построения линий регрессии правомочно с учетом следующих соображений.

Обсуждаемая задача в первую очередь состоит в моделировании и описании зависимости изменения содержаний K_2O в ряду дифференциации, положение в котором фиксируется значением параметра A_c , от основных разновидностей ($A_c = 0,6-0,8$) до кислых ($A_c = 1,4-1,6$). Причем в данном случае зависимость содержаний K_2O от значений A_c в конкретной разновидности пород не рассматривается. Следовательно, в данной задаче нас интересует практически только зависимость условных средних подсовокупностей (разновидностей пород) от значений параметра A_c ,

ее прямолинейный характер и градиент изменения (угол наклона линии регрессии к оси A_c) в исследуемых магматических сериях.

По результатам построения регрессионных прямых для изучаемых выборок следует отметить четко фиксируемую положительную корреляционную зависимость содержания окиси калия от кислотности пород во всех сериях, близость положения условных средних и даже единичных наблюдений в выборках относительно линий регрессии и одинаковую роль K_2O (функция) и A_c (аргумент) во всех сериях. В качестве иллюстрации охарактеризованных зависимостей на рис. 13 показаны линии регрессии K_2O по A_c для четырех базальтоидных серий.

Ниже приводятся эмпирические уравнения регрессии содержаний K_2O от значений параметра A_c и значений параметра A_c от содержаний K_2O для всех исследуемых базальтоидных серий.

Серия	Уравнения регрессии
Сент-Китс	$K_2O = 0,477 + 0,67 \cdot (A_c - 0,98) \pm 0,10;$ $A_c = 0,98 + 1,48 \cdot (K_2O - 0,47) \pm 0,14;$
Тонга	$K_2O = 0,65 + 0,97 \cdot (A_c - 1,17) \pm 0,22;$ $A_c = 1,17 + 1,03 \cdot (K_2O - 0,65) \pm 0,21;$
Лассен-Парк	$K_2O = 1,84 + 2,41 \cdot (A_c - 1,18) \pm 0,26;$ $A_c = 1,18 + 0,41 \cdot (K_2O - 1,84) \pm 0,11;$
Бердс-Ривер	$K_2O = 1,38 + 3,73 \cdot (A_c - 0,92) \pm 0,30;$ $A_c = 0,92 + 0,27 \cdot (K_2O - 1,38) \pm 0,08;$
Шотландия	$K_2O = 1,89 + 4,29 \cdot (A_c - 0,93) \pm 0,41;$ $A_c = 0,93 + 0,23 \cdot (K_2O - 1,89) \pm 0,10;$
Анды чилийские	$K_2O = 1,80 + 3,58 \cdot (A_c - 0,98) \pm 0,64;$ $A_c = 0,98 + 0,23 \cdot (K_2O - 1,80) \pm 0,18;$
Анды перуанские	$K_2O = 2,63 + 3,48 \cdot (A_c - 1,06) \pm 0,38;$ $A_c = 1,06 + 0,29 \cdot (K_2O - 2,63) \pm 0,11;$
Восточная Ирландия	$K_2O = 2,11 + 3,78 \cdot (A_c - 0,89) \pm 1,09;$ $A_c = 0,89 + 0,27 \cdot (K_2O - 2,11) \pm 0,29;$
Исландия	$K_2O = 1,13 + 2,76 \cdot (A_c - 0,92) \pm 0,29;$ $A_c = 0,92 + 0,36 \cdot (K_2O - 1,13) \pm 0,11;$
Бойна	$K_2O = 2,44 + 3,98 \cdot (A_c - 1,01) \pm 0,64;$ $A_c = 1,01 + 0,25 \cdot (K_2O - 2,44) \pm 0,16;$
Пасха	$K_2O = 1,31 + 3,74 \cdot (A_c - 0,91) \pm 0,32;$ $A_c = 0,91 + 0,27 \cdot (K_2O - 1,31) \pm 0,09;$
Кения	$K_2O = 2,30 + 10,61 \cdot (A_c - 0,77) \pm 0,42;$ $A_c = 0,77 + 0,09 \cdot (K_2O - 2,30) \pm 0,04;$
Джебел-Абиад	$K_2O = 3,02 + 10,68 \cdot (A_c - 0,82) \pm 0,70;$ $A_c = 0,82 + 0,09 \cdot (K_2O - 3,02) \pm 0,07;$
Святая Елена	$K_2O = 2,64 + 8,41 \cdot (A_c - 0,78) \pm 0,80;$ $A_c = 0,78 + 0,12 \cdot (K_2O - 2,64) \pm 0,09;$
Восточно-Африканский Рифт	$K_2O = 2,01 + 8,48 \cdot (A_c - 0,71) \pm 0,18;$ $A_c = 0,71 + 0,12 \cdot (K_2O - 2,01) \pm 0,02.$

Проведенный предварительный статистический анализ выборочных совокупностей, представляющих изучаемые базальтоидные серии, и результаты вычисления уравнений регрессии свидетельствуют, что базальтоидные серии с близким по размаху интервалом значений параметра A_c , т.е. сопоставимые в принятых предположениях по степени дифференцированности, могут существенно отличаться как средним (по серии) содержанием K_2O , так и градиентом его изменения. Можно предположить, что изучаемые серии различаются между собой особенностями поведения K_2O в процессе дифференциации. Отсюда можно сформулировать задачу выделения групп базальтоидных серий, близких по поведению K_2O в процессе дифференциации исходного магматического расплава.

Вполне естественно предположить, что построение линий регрессии содержаний K_2O для всех базальтоидных серий на одной двумерной диаграмме с единым масштабом изображения значений аргумента A_c позволит произвести группировку всех базальтоидных серий по характеру изменения содержаний K_2O . Как следует из предшествующего изложения, подобные группы серий с близким градиентом изменения содержаний K_2O должны иметь сопоставимый угол наклона линий регрессии к оси аргумента. Процедура построения линий регрессии для каждой из серий в общем поле регрессии реализуется следующим образом.

Положение линий регрессии для зависимости K_2O от A_c в общем поле регрессии осуществляется с использованием полученных ранее результатов статистического и регрессионного анализов. С этой целью для каждого признака (x и y) вычисляется оценка генерального среднего. При выборе оценки возникает вопрос: какую оценку генерального среднего более рационально использовать в данной задаче: средневзвешенную (с учетом объема выборок):

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots + \bar{x}_i n_i + \dots + \bar{x}_m n_m}{N},$$

где \bar{x}_i – оценка среднего значения x для i -й выборки ($i = 1, 2, \dots, m$), n_i – число наблюдений в i -й выборке, а N – общее число наблюдений по всем выборкам: $N = \sum_{i=1}^m n_i$ и соответственно

$$\bar{y} = \frac{\bar{y}_1 n_1 + \bar{y}_2 n_2 + \dots + \bar{y}_i n_i + \dots + \bar{y}_m n_m}{N}$$

или среднестатистическую оценку генерального среднего, определяемую по оценкам средних значений всех выборок:

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_i + \dots + \bar{x}_m}{m}; \quad \bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \dots + \bar{y}_i + \dots + \bar{y}_m}{m},$$

где m – число выборок.

Если известно, что геологические объекты опробованы равномерно и с одинаковой степенью детальности, то, естественно, более рациональным является применение средневзвешенной оценки генерального среднего. Если же степень опробованности неравноценна или (что чаще всего имеет

Градиенты изменения K_2O относительно A_c и углы наклона эмпирических линий регрессии в общем поле зависимости K_2O-A_c

Серия базальтоидных пород	Градиент изменения K_2O	Угол наклона линии регрессии, град	Серия базальтоидных пород	Градиент изменения K_2O	Угол наклона линии регрессии, град
Сент-Китс	0,34	18,9	Восточная Ирландия	1,93	62,6
Тонга	0,49	26,3	Бойна	2,03	63,8
Лассен-Парк	1,23	50,0	Шотландия	2,19	64,3
Исландия	1,41	54,6	Святая Елена	4,29	76,9
Анды перуанские	1,78	60,6	Восточно-Африканский Рифт	4,33	77,0
Анды чилийские	1,83	61,3	Кения	5,41	79,5
Бердс-Ривер	1,90	62,3	Джебел-Абиад	5,45	79,6
Пасха	1,91	62,3			

место) неизвестна, то применение среднестатистической оценки генерального среднего более надежно, так как она нивелирует разницу в степени опробованности или, что в нашем случае равноценно, различие в объеме выборочных совокупностей. Именно по этой причине в рассматриваемом случае использовалась среднестатистическая оценка генерального среднего.

Среднестатистическая оценка генерального среднего для показателя общей кислотности пород A_c для 15 рассматриваемых базальтоидных серий: $A_c = (0,98 + 1,17 + 1,18 + 0,93 + 0,92 + 0,98 + 1,06 + 0,89 + 0,92 + 1,01 + 0,91 + 0,77 + 0,82 + 0,78 + 0,71) : 15 = 0,94$.

Аналогичная оценка генерального среднего для содержаний K_2O в исследуемой совокупности серий: $K_2O = (0,47 + 0,65 + 1,84 + 1,89 + 1,38 + 1,80 + 2,61 + 2,11 + 1,13 + 2,44 + 1,13 + 2,30 + 3,02 + 2,64 + 2,01) : 15 = 1,84\%$.

Эмпирический коэффициент пропорциональности для зависимости K_2O от A_c для объединенной выборки из всех серий: $K_{K_2O} = 1,84 : 0,94 = 1,96$.

С помощью этого обобщенного по всем выборкам коэффициента пропорциональности была рассчитана шкала для A_c при графическом изображении зависимости K_2O-A_c на общем графике.

Имея выборочные значения генеральных средних K_2O и A_c для всех базальтоидных серий и размах значений содержаний K_2O и величины A_c для каждой серии, можно вычислить градиенты изменения K_2O относительно A_c (коэффициенты диспропорциональности) для каждой серии в общем поле зависимости по формуле

$$D_{K_2O} = \frac{\tilde{K}_{K_2O}}{\hat{K}_{K_2O}} = \frac{\frac{\sigma_{K_2O}}{\sigma_{A_c}}}{\frac{K_2O}{A_c}}$$

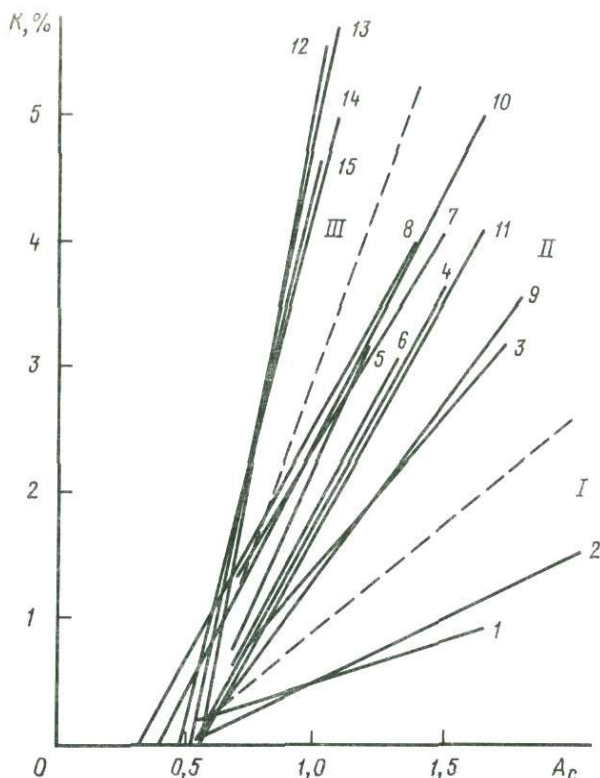


Рис. 14. Результаты группирования дифференцированных базальтоидных серий на основе линий регрессии $K-A_c$

I — толеитовая группа: 1 — Сент-Китс, 2 — Тонга; *II* — известково-щелочная группа: 3 — Лассен-Пик, 4 — Исландия, 5 — Анды (Чили), 6 — Анды (Перу), 7 — Бердс-Ривер, 8 — Пасха, 9 — Восточная Ирландия, 10 — Бойна, 11 — Шотландия; *III* — щелочная группа: 12 — Святая Елена, 13 — Восточно-Африканский щит, 14 — Кения, 15 — Джебел-Абиад

Градиенты изменения содержаний K_2O (тангенсы углов наклона) и вычисленные по ним углы наклона эмпирических линий регрессии каждой серии в общем поле зависимости K_2O от A_c даны в табл. 20.

Из рассмотрения табл. 20 и рис. 14 нетрудно заметить, что общая выборка дифференцированных базальтоидных серий по градиентам изменения содержаний K_2O относительно A_c разделилась на три группы (*I*–*III*) в полном соответствии с известными представлениями [6, 7] о типах базальтоидного магматизма.

I. Низкокалийевые, преимущественно толеитовые, серии островных дуг — Сент-Китс и Тонга. В этих сериях диапазон изменения содержаний K_2O составляет не более 0,3–1,3% — по верхней границе низкокалийевых пород; углы наклона линий регрессии K_2O от A_c в общем поле зависимости не превышает 30° .

II. Известково-щелочные (нормальнокалийевые) серии. Основная масса пород в них (от андезитов до дацитов и риолитов) отличается более высо-

кими содержаниями K_2O (до 3%). Углы наклона линий регрессии K_2O от A_c в общем поле зависимости K_2O-A_c находятся в пределах $50-65^\circ$. В эту группу вошли серии: Лассен-Парк, Исландия, Анды чилийские, Анды перуанские, Бердс-Ривер, Пасха (серия субщелочных, но преимущественно натриевых базальтов), Восточная Ирландия, Бойна (субщелочная натриевая серия), Шотландия.

III. Щелочно-базальтовая группа с сериями океанических островов – Святая Елена и рифтовых зон – Восточно-Африканский Рифт, Кения и Джебел-Абиад. Для этой группы серий характерен самый высокий градиент изменения содержаний K_2O с увеличением A_c и вследствие этого самые большие углы наклона линий регрессии K_2O от A_c в общем поле зависимости (более 75°). Очевидно, что градиент изменения K_2O является отражением типичного для данной группы серий ряда дифференциатов, соответствующего тренду кристаллизационной дифференциации от щелочных базальтов до трахиандезитов.

Представляет интерес использование результатов исследования зависимости K_2O от значений A_c в группах базальтоидных серий для решения такого вопроса: имеют ли все серии общую точку пересечения линий регрессии? Наличие такой точки можно интерпретировать как начало магматической дифференциации при формировании серий из одного источника (по составу, а не по месту!) независимо друг от друга. Альтернативное предположение: дифференциация происходит последовательно – из малокалиевого субстрата выделяется среднекалиевый магматический материал, а из среднекалиевого – высококалиевый.

Проверка справедливости одного из приведенных выше альтернативных предположений осуществлялась на базе построения регрессионных прямых K_2O для базальтоидных серий, принадлежащих разным типам базальтоидного магматизма. На рис. 14 приведены регрессионные прямые изученных базальтоидных серий, разделенных по углу наклона линии регрессии на три группы, соответствующие магмам толеитового (I группа – две серии), известково-щелочного (II группа – 9 серий) и щелочно-базальтового (III группа – 4 серии) типов. Изображенные на рис. 14 регрессионные прямые по углу наклона к оси x достаточно уверенно подразделяются на три указанные выше группы. В связи с тем что приведенные линии регрессии являются обобщением данных конкретных выборочных совокупностей (часто небольшого объема), очевидно, наклон каждой конкретной линии обусловлен не только природными закономерностями (дифференциацией), но и влиянием случайных причин. В связи с этим при проверке правомочности одного из двух обсуждаемых предположений для снижения влияния случайных факторов было осуществлено обобщение фактического материала по градиенту изменения содержаний K_2O в процессе дифференциации (A_c) для серий каждого типа. Для этого по каждой группе серий были определены средние параметры: средний градиент изменения K_2O от A_c , среднее значение A_c и среднее содержание K_2O . Результаты этих вычислений приведены в табл. 21. Затем на графиках в координатах A_c (аргумент) – K_2O (функция) были нанесены точки, соответствующие средним значениям признаков (A_c и K_2O) для каждой группы серий. Через эти точки в соответствии со средним значением градиента изменения K_2O от A_c

Эмпирические значения параметров изменения содержаний K_2O от значений A_c в разных типах базальтоидного магматизма

Тип базальтоидного магматизма	Средний градиент изменения K_2O от A_c	Среднее значение A_c	Среднее содержание K_2O (%)
Толеитовый (I группа)	0,42	1,08	0,56
Известково-щелочной (II группа)	1,75	0,98	1,83
Щелочно-базальтовый (III группа)	4,79	0,77	2,49

были проведены соответствующие прямые, которые по предположению отражают типовое (по изученным материалам) для каждого типа базальтового магматизма изменение содержаний K_2O в последовательности продуктов дифференциации. Из этих построений и приведенных выше предположений явно следует, что точка пересечения таких прямых (если она существует), во-первых, соответствует гипотетическому составу исходной магмы (по содержаниям K_2O и значениям A_c) и, во-вторых, подтверждает гипотезу о независимом развитии процессов дифференциации каждого типа из одной исходной магмы. Отсутствие такой точки пересечения всех трех прямых, т.е. последовательное причленение прямых каждого из типов друг к другу, служило бы косвенным подтверждением гипотезы о последовательном развитии очага и изменении состава исходной магмы от одного типа к другому и далее — к третьему.

На рис. 15 приведены результаты построения усредненных по группам трендов изменения содержаний K_2O от A_c в сериях трех типов. Представленная картина соотношения трендов поведения K_2O в процессе дифференциации базальтовых серий трех типов дает основание для следующих выводов. Из приведенного графика вполне очевидно, во-первых, что предположение о независимости развития всех трех типов базальтов из исходного магматического материала одного состава не подтверждается. Естественно, что с большим основанием может быть признано по наблюдаемым фактам второе из альтернативных предположений, однако с некоторыми коррективами.

С позиций сформулированной выше задачи и принятых предположений, точка "А" на графике (см. рис. 15), координаты которой — $A_c = 0$; $K_2O = 0,22\%$, соответствует составу гипотетического исходного материала, исходного для трендов дифференциации базальтоидных серий двух типов: толеитового и известково-щелочного. Причем наблюдаемое на графике соотношение трендов серий этих двух типов свидетельствует в пользу предположения о независимом развитии процессов дифференциации двух типов на базе исходного материала предположительно одинакового состава по K_2O и по значениям A_c .

С учетом всех принятых предположений базальтоидные серии щелочно-базальтового типа следует считать дериватом процесса развития известково-щелочного магматизма. Исходным материалом для базальтоидных серий щелочно-базальтового типа можно предполагать магматический материал

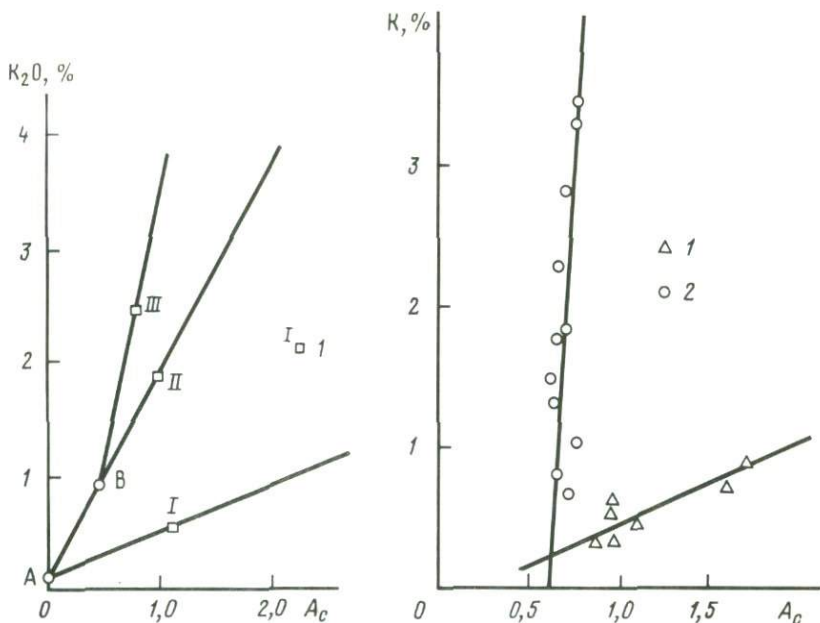


Рис. 15. Обобщенные линии зависимости K_2O-A_c для групп базальтоидных серий I — толеитовая; II — известково-щелочная; III — щелочная; I — двумерные средние для соответствующих групп

Рис. 16. Толеитовая (1) и щелочная (2) серии базальтоидных пород островов Фиджи

известково-щелочного состава со средним значением $A_c = 0,43$ и содержанием $K_2O = 0,86\%$.

При интерпретации рассматриваемых результатов необходимо учитывать некоторые немаловажные обстоятельства. Пути дифференциации при базальтоидном магматизме нами оценивались только с позиций характера изменения содержаний K_2O от значений показателя дифференциации A_c , причем тренд содержаний K_2O по A_c аппроксимировался моделью прямолинейной зависимости. Кроме того, как ни интересны приведенные петрологические выводы, основной целью примера все же явилась демонстрация возможностей и продуктивности применения предлагаемой методики регрессионного анализа к обоснованию выводов при решении сложных и разнообразных задач геологии. Вместе с тем результаты анализа зависимости K_2O-A_c на базе построения линий ортогональной регрессии для набора объектов (дифференцированных базальтоидных серий) дают основание для следующих петрологических выводов [6, 7].

1. Для главных типов дифференцированных базальтоидных серий из континентальных и океанических провинций выявлены линейные тренды возрастания и концентраций K_2O в зависимости от увеличения общей кислотности магматических расплавов (показатель дифференциации A_c).

2. По особенностям закономерного изменения концентраций K_2O выделяются три явно обособленные группы базальтоидных серий, различаю-

Содержания K_2O (в %) и значения общего показателя кислотности A_c в породах базальтоидной серии Фиджи

Номер п/п	Номер пробы	A_c	K_2O , %	Номер п/п	Номер пробы	A_c	K_2O , %
1	6095	0,61	1,48	10	6102	0,77	3,30
2	6096	0,63	1,31	11	1601	0,77	3,46
3	6099	0,66	1,78	12	6082	0,86	0,34
4	6079	0,67	0,81	13	6083	0,95	0,57
5	6097	0,67	2,29	14	6084	0,95	0,66
6	6080	0,70	0,66	15	6086	0,98	0,37
7	6098	0,70	1,80	16	6085	1,10	0,46
8	6100	0,72	2,79	17	6087	1,59	0,75
9	6081	0,75	1,03	18	6088	1,70	0,93

щихся как уровнем содержаний K_2O , так и градиентом его изменения относительно показателя дифференциации A_c — наименьшим для группы толеитовых серий (Сент-Китс и Тонга) и наибольшим для группы щелочно-базальтовых серий (Кения, Джебел-Абиад и др.).

3. Усредненные линейные тренды изменения концентраций K_2O , выявленные для главных типов базальтоидного магматизма, позволяют предположить независимое развитие из исходного материала одного состава (практически пикритового) для серий двух типов (толеитового и известково-щелочного) и рассматривать щелочно-базальтовый магматизм как дериват известково-щелочной магмы.

Регрессионный анализ зависимости K_2O-A_c в базальтоидных сериях позволяет выявить еще один важный аспект геохимических особенностей природных объектов. Рассмотрим его на примере серии Фиджи, не участвовавшей в общем анализе базальтоидных серий. В табл. 22 представлены исходные данные по содержаниям K_2O и значениям A_c в общей совокупности наблюдений, а на рис. 16 — положение точек наблюдения на диаграмме K_2O-A_c . Из рис. 16 отчетливо видно, что серия Фиджи является существенно неоднородной по градиенту возрастания K_2O с увеличением A_c , а состоит из двух резко различных серий пород:

1-я серия — Фиджи-1 (№ 6082—6088) по параметрам зависимости K_2O-A_c (табл. 23) относится к низкокалиевой, толеитовой группе серий (как и Сент-Китс, Тонга) — угол наклона линии регрессии в общем (для всех базальтоидных серий) поле зависимости K_2O-A_c составляет всего $25,1^\circ$.

2-я серия — Фиджи-2 (№ 6079—6081, 6095—6102) — по параметрам зависимости K_2O-A_c (см. табл. 23) относится к щелочно-базальтовой группе серий (как и серии Кения, Джебел-Абиад и др.), причем угол наклона линии регрессии к оси абсцисс в общем поле зависимости K_2O-A_c очень большой — $85,6^\circ$.

Полученные результаты исследования зависимости содержаний K_2O (индикаторного компонента для типов базальтоидного магматизма) от значений показателя общей кислотности магм (A_c) с достаточным основа-

Таблица 23

Оценки параметров зависимости содержаний K_2O (в %) и значений показателя общей кислотности A_c в базальтовых сериях Фиджи

Статистические оценки	Символы	Фиджи-1		Фиджи-2	
		A_c	K_2O	A_c	K_2O
Число наблюдений	n	7		11	
Среднее значение	\bar{x}, \bar{y}	1,16	0,55	0,70	1,88
Стандартное отклонение	σ_s	0,34	0,24	0,05	0,97
Коэффициент пропорциональности	\tilde{K}	2,094	0,448	0,369	2,709
Коэффициент регрессии	\tilde{K}	1,393	0,718	0,056	1,793
Стандартное отклонение точек наблюдения от линии регрессии	$\tilde{\sigma}$	0,24	0,17	0,05	0,92
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,22$	$\pm 0,16$	$\pm 0,03$	$\pm 0,62$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{\Pi}_{\text{экс}}$	$+0,143$		$+0,273$	
Толерантные пределы отклонений точек наблюдения от линии регрессии	\tilde{T}	$\pm 0,59$	$\pm 0,42$	$\pm 0,11$	$\pm 2,05$
Коэффициент диспропорциональности	D	0,665	1,503	0,151	6,612
Роль переменных		Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс на частной диаграмме	α_r	$56,4^\circ$		$81,4^\circ$	
Угол наклона линии регрессии к оси абсцисс на общей диаграмме	$\alpha_{об}$	$28,8^\circ$		$85,6^\circ$	
Коэффициент корреляции поля регрессии	$R_{x,y}$	$+0,652$		$\pm 0,526$	
Критическое значение коэффициента корреляции	$R_{0,05}$	0,930		0,663	

нием позволяют предположить, что разновидности пород магматических объектов, относящиеся к единому комплексу (единой серии дифференциатов), характеризуются единым градиентом изменения петрохимических характеристик.

При исследовании зависимости двух признаков во множестве геологических объектов может встретиться вариант, отличающийся от выше-рассмотренного. Градиент изменения значений признаков в объектах может быть сопоставимым, но общая совокупность объектов может быть неоднородной по отношению к общей линии зависимости признаков, т.е. могут существовать частные выборки с параллельными и субпараллельными линиями регрессии. Исследование зависимости значений признаков с помощью метода ортогональной регрессии позволяет решать задачу разделения таких неоднородных совокупностей, что будет показано на примере изучения зависимости содержания серы и меди в серноколчеданных рудах уральских колчеданных месторождений.

Серный колчедан (пириты) составляют основу рудной массы колчеданного оруденения. Так например, в уральских месторождениях пирит составляет 86% от всей сульфидной массы [18]. Вследствие этого весьма существенным для выявления условий колчеданного рудообразования является определение структуры и параметров серноколчеданного оруденения.

Как показали предшествующие исследования [14, 15], степень метаморфизма серноколчеданных руд соответствует степени метаморфизма вмещающих вулканогенных пород при линзовидной форме серноколчеданных руд и согласном или субсогласном их залегании по отношению к вмещающим породам. Отсутствуют указания хотя бы на один случай пересечения серноколчеданными рудами субвулканических тел, даек и малых интрузий диоритового, гранодиоритового или иного состава. В то же время различные, в том числе и секущие, взаимоотношения медь- и цинк-содержащих руд с такими же субвулканическими телами, дайками и малыми интрузиями наблюдается.

Перечисленные особенности серноколчеданного оруденения имеют логическое объяснение только при допущении, что самостоятельные серноколчеданные рудные тела, месторождения, а вместе с ними в вообще серноколчеданная составляющая колчеданного оруденения Урала являются сингенетичными по отношению к вмещающим вулканогенным и вулканогенно-осадочным комплексам, т.е. являются эксгальционно-гидротермально-осадочными, в то время как медная и полиметаллическая составляющие оруденения обусловлены более поздними процессами и являются наложенными на уже сформировавшиеся к этому времени серноколчеданные руды. Если такое предположение о одновременности двух процессов рудообразования верно, то вполне естественно предположить, что имеются следы переработки первичных серноколчеданных руд — протопиритов [12] более поздними процессами формирования медной и полиметаллической составляющих. Применение регрессионного анализа для выявления особенностей взаимного поведения серы и меди в серноколчеданных рудах позволило обнаружить следы такой переработки.

В массивных серноколчеданных рудах в тех или иных количествах присутствует весь комплекс химических элементов, присущих колчедан-

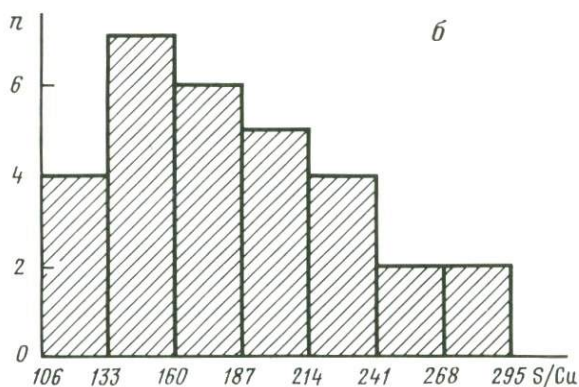
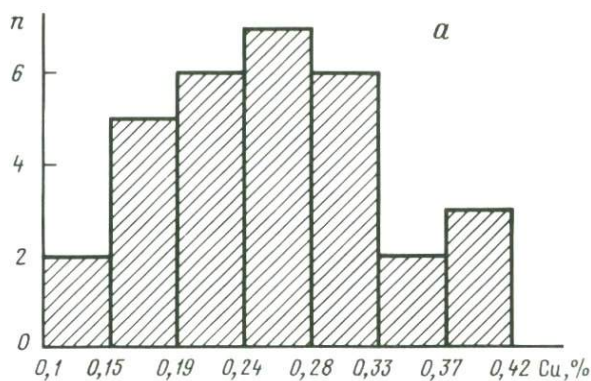


Рис. 17. Гистограмма распределения содержаний меди (а) и значений отношения S/Cu (б) в серноколчеданных рудах

ному оруденению Урала вообще [18], причем такие металлы, как медь, цинк, свинец и барий, обычно образующие собственные рудные скопления, в серноколчеданных рудах присутствуют как элементы-примеси наряду с селеном, теллуrom, кобальтом, мышьяком, висмутом, кадмием, индием и др. В этом типе руд главными, рудообразующими являются только железо и сера. Из обширного набора химических элементов, присутствующих в серноколчеданных рудах, для дальнейших исследований нами были выбраны для исследований сера и медь. Сера — как главный, рудообразующий элемент, а медь — как индикатор процессов и изменений, которым подвергалось наблюдаемое в настоящее время оруденение.

Применение регрессионного анализа содержаний серы и меди осуществлялось для получения ответов на следующие вопросы. 1. Существует ли какая-либо линейная зависимость между содержаниями серы и меди первого уровня концентрации в массивных серноколчеданных рудах (следует учитывать, что, как было показано ранее [18], между серой и медью

Регрессионный анализ содержаний меди и серы
в массивных серноколчеданных рудах Урала

Номер п/п	Сера, %	Медь, %	$\tilde{\Delta}\text{Cu}$	S/Cu	Номер п/п	Сера, %	Медь, %	$\tilde{\Delta}\text{Cu}$	S/Cu
1	40,36	0,27	0,130	149	17	44,60	0,26	0,004	172
2	40,50	0,18	0,035	225	18	44,60	0,34	0,074	131
3	40,74	0,25	0,099	163	19	45,30	0,28	0,005	160
4	41,50	0,15	0,020	277	20	46,00	0,10	-0,194	460
5	41,84	0,23	0,049	182	21	46,50	0,32	0,013	145
6	41,95	0,17	-0,014	242	22	46,67	0,31	-0,002	151
7	42,10	0,39	0,202	108	23	46,97	0,42	0,100	112
8	42,15	0,18	-0,010	234	24	47,19	0,16	-0,166	295
9	42,68	0,31	0,106	138	25	47,20	0,22	-0,106	215
10	42,55	0,17	-0,030	250	26	47,50	0,24	-0,095	198
11	42,70	0,30	0,096	142	27	48,50	0,25	-0,112	194
12	43,19	0,32	0,102	135	28	48,60	0,26	-0,104	187
13	43,35	0,20	-0,022	217	29	49,00	0,22	-0,155	333
14	43,70	0,21	-0,022	208	30	50,00	0,35	-0,052	143
15	44,40	0,26	0,009	171	31	50,38	0,29	-0,123	174
16	44,49	0,42	0,167	106					

второго уровня концентрации, образующей основные промышленные типы руд, никакой зависимости не наблюдается)? 2. Какова роль этих химических элементов в их взаимоотношениях? 3. Представляют ли серноколчеданные руды однородную группу по содержаниям меди и по взаимоотношению сера—медь, а если нет, то как выделить однородные (по указанным признакам) группы руд и чем они различаются?

Исходные данные по содержаниям серы и меди в серноколчеданных рудах уральских колчеданных месторождений даны в табл. 24 (графы 2 и 3), там же приведены результаты регрессионного анализа содержаний меди от содержаний серы (графа 4) и вычисленные для всех объектов соотношения S/Cu (графа 5).

Как по содержаниям меди (рис. 17, а), так и по величине значений S/Cu отношений (рис. 17, б) общая выборка серноколчеданных руд по использованному в методике показателю оказалась однородной, т.е. ни содержания меди, ни величины отношений сера—медь не указывают на какую-либо явную неоднородность в изменениях изучаемых признаков серноколчеданных руд. Вычисленные параметры распределения содержаний этих элементов (табл. 25) свидетельствуют о том, что эмпирические распределения содержаний и серы, и меди симметричны и не противоречат нормальной модели распределения величин.

Вычисление параметров для описания зависимости содержаний серы и меди (табл. 26) позволило установить, что роль этих элементов в их взаимоотношениях неодинакова. Сера устойчиво выступает в роли аргумента ($D_S = 0,21$), а медь — в роли функции ($D_{Cu} = 4,76$). Для содер-

Параметры распределения содержаний серы и меди
в массивных серноколчеданных рудах Урала

Статистические оценки	Символы	Сера, %	Медь, %
Число наблюдений	n	31	31
Среднее значение	\bar{x}, \bar{y}	44,75	0,26
Стандартное отклонение	σ	2,90	0,08
Доверительные интервал для среднего значения	λ	$\pm 1,06$	$\pm 0,03$
Показатель асимметрии распределения	$\tilde{\Pi}_{ас}$	-0,160	-0,065
Показатель эксцесса распределения	$\tilde{\Pi}_{экс}$	0,335	-0,032
Толерантные пределы значений при $(1 - q) = 0,90$			
а) минимальные	T_{\min}	38,83	0,10
б) максимальные	T_{\max}	50,67	0,42

Т а б л и ц а 26

Параметры зависимости содержаний меди от содержаний серы
в массивных серноколчеданных рудах Урала

Статистические оценки	Символы	Сера, %	Медь, %
Число наблюдений	n	31	
Среднее значение	\bar{x}, \bar{y}	44,75	0,26
Стандартное отклонение	σ	2,90	0,08
Коэффициент пропорциональности	\tilde{K}	172,0	0,0058
Коэффициент регрессии	\tilde{k}	36,25	0,0276
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	3,63	0,10
Доверительный интервал линии регрессии при $(1 - q) = 0,90$	λ	$\pm 1,33$	$\pm 0,04$
Показатель эксцесса распределения относительно регрессии	$\tilde{\Pi}_{экс}$	+0,226	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 7,41$	$\pm 0,204$
Коэффициент диспропорциональности	D	0,21	4,76
Роль переменных		Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	78,1°	
Коэффициент корреляции поля регрессии	$R_{x, y}$	+0,375	
Критическое значение коэффициента корреляции при $(1 - q) = 0,90$	$R_{\text{крит}}$	0,366	
Зависимость		Прямая	

жаний серы и меди фиксируется хотя и довольно слабая, но все же значимая положительная корреляционная зависимость: $R_{x, y} = +0,375 > R_{\text{крит}} = 0,366$.

Однако по результатам построения поля и линии регрессии выявилась одна любопытная деталь. Будучи однородной по содержаниям меди и по величинам S/Cu отношений, общая выборка серноколчеданных руд ока-

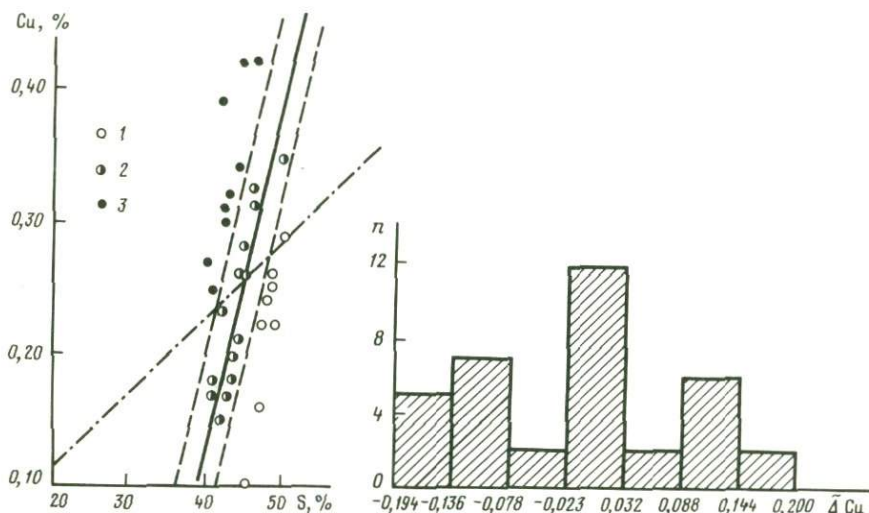


Рис. 18. Группирование серноколчеданных руд по содержанию меди:

1-3 - руды: 1 - маломедистые; 2 - среднемедистые; 3 - высокомедистые

Рис. 19. Гистограмма распределения отклонений содержаний меди от рассчитанных по уравнению регрессии (ΔCu)

залась неоднородной по величине ΔCu . Даже при первом взгляде на поле регрессионной зависимости (рис. 18) можно заметить, что точки наблюдения расположены не хаотично, а вытянуты цепочками субпараллельно линии регрессии. Следовательно, наиболее надежная дискриминация общей совокупности может быть осуществлена в сечении признакового пространства в направлении, перпендикулярном вытянутости роя точек, т.е. относительно общей линии регрессии. Именно по этому признаку (по величине отклонений точек наблюдения от общей линии зависимости, табл. 24, графа 4) и была проведена проверка выборки сернокислотных руд на однородность. Результаты проверки (рис. 19) подтвердили эту неоднородность. На рисунке по оси ординат откладывается число наблюдений, а по оси абсцисс - величина отклонения наблюдаемых конкретных точек от рассчитанных с помощью уравнения регрессии для заданных содержаний серы, т.е. от линии регрессии.

Таким образом, по величине отклонений реальных содержаний меди от расчетных общую выборку серноколчеданных руд уральских колчеданных месторождений правомочно разделить на три составляющие подсовокупности (см. табл. 24).

1. Серноколчеданные руды, составы которых расположены на графике вблизи общей линии зависимости (линии регрессии) - анализы 2, 4-6, 8, 10, 13-15, 17, 19, 21, 22 и 30. Для этих руд величина ΔCu колеблется в пределах от $-0,050$ до $+0,060$.

2. Серноколчеданные руды, точки наблюдения которых расположены выше линии регрессии по оси ординат, т.е. эти руды относительно обогащены медью - $(+0,060 < \Delta \text{Cu} < +0,202)$; анализы 1, 3, 7, 7, 11, 12, 16, 18, 23.

3. Серноколчеданные руды, точки наблюдения которых локализируются ниже линии регрессии (по оси ординат). Это наиболее чистые, серноколчеданные руды, с относительно низкими содержаниями меди ($-0,190 < \tilde{\Delta}Cu < -0,050$); анализы 20, 24–29, 31.

Одна из причин, почему разделение общей выборки серноколчеданных руд не выявилось непосредственно по содержаниям меди, рассмотрена выше. Здесь имеет смысл не только остановиться на вопросе, почему эта неоднородность не выявлена и по величинам S/Cu отношений, но и рассмотреть причины, которые обеспечили возможность выявления неоднородности общей совокупности при регрессионном анализе. Такая возможность прямо следует из особенностей предлагаемой методики. На рис. 19 поле регрессии изображено в пропорциональном масштабе, т.е. с использованием метода рационального расчета шкал для построения поля регрессии. Нетрудно заметить, что ось поля регрессии наклонена под значительным углом к линии пропорциональной зависимости ($D_{Cu} = 4,76$, что соответствует углу наклона линии регрессии к оси абсцисс $78,1^\circ$). Вследствие такого значительного отклонения линии регрессии от линии пропорциональности (проходящей, как известно, под углом 45° к координатным осям) происходит взаимное перекрытие групп, выделяемых по $\tilde{\Delta}Cu$, значениями S/Cu отношений.

Покажем это на рассматриваемом примере. Действительно, если взять группу руд с самыми низкими содержаниями меди, то S/Cu отношение для этой группы меняется от 174 (ан. 31) до 460 (ан. 20); в группе руд со средними содержаниями меди такой показатель меняется от 143 (ан. 30) до 277 (ан. 4), а в группе руд с высокими содержаниями меди S/Cu отношение меняется от 106 (ан. 16) до 163 (ан. 3). Из этого следует, что S/Cu отношение не может быть четким показателем принадлежности серноколчеданных руд к той или иной группе по относительному обогащению их медью, так как изменение соотношения сера–медь происходит не по пропорциональной схеме, а более сложным путем, о чем может свидетельствовать и различие в величине коэффициентов вариации

$$V_S = \sigma_S / \bar{S} = 0,065; \quad V_{Cu} = \sigma_{Cu} / Cu = 0,308.$$

Таким образом, регрессионный анализ зависимости содержания серы и меди в серноколчеданных рудах уральских колчеданных месторождений по предлагаемой методике показал, что эти руды нельзя рассматривать как геохимически однородную группу руд: их необходимо разделить на три группы по относительным содержаниям меди. Неоднородность совокупности серноколчеданных руд может служить основанием для двух альтернативных предположений: эти группы имеют разное происхождение или эти группы образовались в результате последующих изменений генетически однородной группы серноколчеданных руд – протопиритов.

Рассмотрим более подробно, чем же отличаются друг от друга выделенные группы серноколчеданных руд? Первая (маломедистая) и вторая (среднемедистая) группы руд близки по средним содержаниям меди (0,22% и 0,23% соответственно, $t = 0,34 < t_{0,05}(22) = 2,09$), но значительно различаются содержаниями серы – 48,05% и 43,95% соответственно

($t = 4,95 > t_{0,05} (22) = 2,09$). Вторая и третья (высокомедистая) группы серноколчеданных руд близки между собой по средним содержаниям серы (43,95% и 43,09%, $t = 0,87 < t_{0,05} (21) = 2,09$), но резко различаются средними содержаниями меди – 0,23% и 0,34% соответственно ($t = 4,29 > t_{0,05} (21) = 2,09$). Первая и третья группы серноколчеданных руд различаются, естественно, как содержаниями серы, так и содержаниями меди.

Вследствие того что общая выборка серноколчеданных руд оказалась неоднородной и состоит, по крайней мере, из трех групп, различающихся параметрами регрессионной зависимости, все построения регрессионного анализа для общей выборки являются неправомерными; значит, регрессионный анализ зависимости содержаний серы и меди необходимо провести отдельно для каждой выделенной группы серноколчеданных руд.

Маломедистые серноколчеданные руды. Объем выборки для данной группы руд является самым меньшим (всего 8 анализов) – либо руды чисто серноколчеданных месторождений (Куль-Юрт-Тауское, Андреевское, Арамилское), либо руды изолированных серноколчеданных тел колчеданных месторождений других геохимических типов [18]. Эти руды наименее изменены и по своим свойствам приближаются к протопиритам [12]. Серноколчеданные руды рассматриваемого типа характеризуются самым высоким содержанием серы – 48,05%. Интересно отметить, что соотношение между медью и цинком в этом типе руд, равное 1,30, полностью соответствует кларковому соотношению этих металлов в изверженных породах основного состава [10, 13]. Параметры распределения содержаний серы и меди в этом типе руд приведены в табл. 27.

Для маломедистой группы серноколчеданных руд был проведен регрессивный анализ зависимости содержаний меди от содержаний серы, результаты которого представлены в табл. 28. Уравнение регрессии для описания зависимости содержаний меди от содержаний серы в маломедистых рудах имеет следующий вид: $Cu = 0,22\% + 0,044 \cdot (S - 48,05\%) \pm 0,07\%$.

Обращают на себя внимание очень высокий (9,78) коэффициент диспропорции изменения содержаний меди, а также высокое (+0,813) значение оценки коэффициента корреляции между серой и медью. Общий вид поля регрессионной зависимости содержаний меди от содержаний серы в незатронутых изменениями маломедистых серноколчеданных рудах представлен на рис. 20, а.

Среднемедистые серноколчеданные руды. Руды этого типа содержат больше примесей, чем маломедистые. S/Cu отношение здесь сдвигается в пользу меди (183 против 221 в маломедистых). Параметры распределения содержаний серы и меди в среднемедистых рудах были приведены в табл. 27. Данная группа руд объединяет серные колчеданы Юрьевского, Карпушихинского, Гарньерского, Весеннего, Октябрьского, Учалинского, Сибайского и других месторождений.

Результаты регрессионного анализа зависимости содержаний меди от содержаний серы в рудах этого типа (см. табл. 28) показывают, что коэффициент диспропорциональности для меди в рудах значительно ниже (4,42), чем в маломедистых, и для рассматриваемой группы руд эмпирический

Параметры распределения содержаний серы и меди в выделенных по \bar{X} Cu типах серноколчеданных руд

Статистические оценки	Символы	Типы серноколчеданных руд					
		Маломедистые		Среднемедистые		Высокомедистые	
		Сера, %	Медь, %	Сера, %	Медь, %	Сера, %	Медь, %
Число наблюдений	n	8	8	14	14	9	9
Среднее значение	x, y	48,05	0,22	43,93	0,23	43,05	0,34
Стандартное отклонение	σ	1,35	0,06	2,56	0,06	2,05	0,06
Доверительный интервал среднего значения при $(1-q) = 0,90$	λ	$\pm 1,33$	$\pm 0,05$	$\pm 1,48$	$\pm 0,03$	$\pm 1,58$	$\pm 0,05$
Показатель асимметрии распределения	P_{ac}	0,00	0,25	-0,14	-0,07	-0,11	-0,22
Показатель эксцесса распределения	$P_{экс}$	0,25	0,00	0,14	0,29	0,11	0,33
Толерантные пределы значений при $(1-q) = 0,90$							
а) минимальные	T_{min}	44,85	0,08	38,40	0,09	38,35	0,20
б) максимальные	T_{max}	51,25	0,34	49,46	0,35	47,87	0,46

коэффициент корреляции содержаний элементов имеет высокое положительное значение. Содержания меди в среднемедистых серноколчеданных рудах можно рассчитывать по следующему уравнению:

$$Cu = 0,23\% + 0,023 \cdot (S - 43,93\%) \pm 0,05\%.$$

Общий вид поля регрессионной зависимости содержаний меди от содержаний серы можно видеть на рис. 20, б.

Высокомедистые серноколчеданные руды. Наиболее измененные серноколчеданные руды уральских колчеданных месторождений характеризуются самыми высокими содержаниями меди и самыми низкими содержаниями серы, что отражается и на оценке S/Cu отношения — 128. В эту группу вошли анализы руд Молодежного, Маканского, Джусинского, Троицкого и других месторождений. Параметры распределения содержаний меди и серы (см. табл. 27) свидетельствуют о том, что нормальная модель не соответствует эмпирическому распределению содержаний меди в этих рудах ($P_{ac} = -0,222$).

Не менее интересными представляются и результаты регрессионного анализа зависимости содержаний меди от содержаний серы в высокомедистой группе серноколчеданных руд (см. табл. 28). Хотя для меди и сохраняется довольно высокий коэффициент диспропорциональности ($D_{Cu} = 3,72$), для описания зависимости содержаний меди и серы линей-

Параметры зависимости содержаний меди и серы в разных типах (по $\tilde{\Delta}\text{Cu}$) серноколчеданных руд

Статистические оценки	Символы	Типы серноколчеданных руд					
		Маломедистые		Среднемедистые		Высокомедистые	
		Сера, %	Медь, %	Сера, %	Медь, %	Сера, %	Медь, %
Число наблюдений	n	8		14		9	
Среднее значение	\bar{x}, \bar{y}	48,05	0,22	43,95	0,23	43,05	0,34
Стандартное отклонение	σ	1,35	0,06	2,56	0,06	2,05	0,06
Коэффициент пропорциональности	\tilde{K}	221,0	0,0045	191,0	0,0052	128,0	0,0078
Коэффициент регрессии	\tilde{K}	22,13	0,044	42,67	0,023	34,17	0,029
Стандартное отклонение от линии регрессии	σ	0,75	0,03	1,07	0,025	1,37	0,04
Доверительный интервал линии регрессии при $(1 - q) = 0,90$	$\tilde{\lambda}$	$\pm 0,63$	$\pm 0,028$	$\pm 0,62$	$\pm 0,014$	$\pm 1,05$	$\pm 0,031$
Экссесс распределения относительно линии регрессии	$\tilde{\Pi}_{\text{экс}}$	0,500		0,571		0,556	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 1,78$	$\pm 0,08$	$\pm 2,37$	$\pm 0,054$	$\pm 3,16$	$\pm 0,092$
Коэффициент диспропорциональности	D	0,102	9,78	0,227	4,43	0,267	3,72
Роль переменных		Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	$84,2^\circ$		$77,3^\circ$		$75,0^\circ$	
Коэффициент корреляции поля регрессии	$R_{x,y}$	+0,813		+0,792		+0,667	
Критическое значение коэффициента корреляции	$R_{\text{крит}}$	0,800		0,557		0,770	
Зависимость		Прямая		Прямая		Прямая	

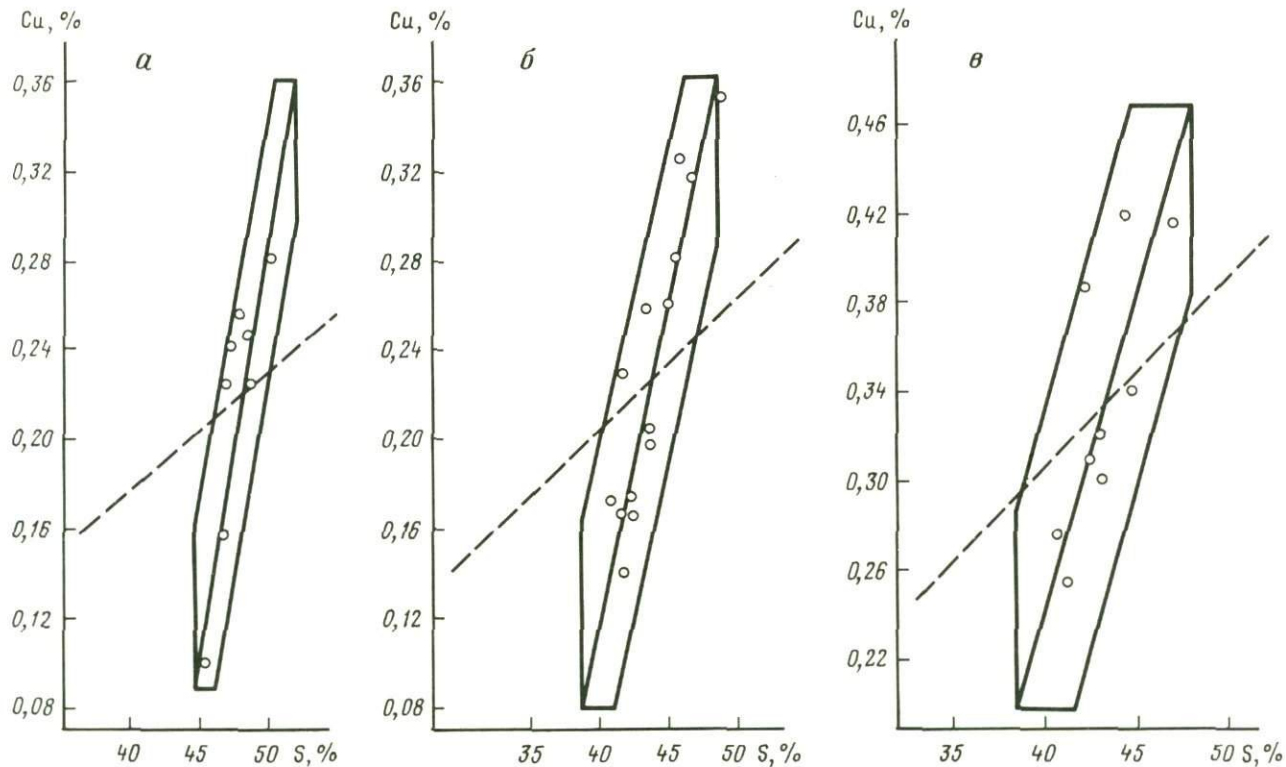


Рис. 20. Поле регрессии для содержаний меди и серы в различных типах серноколчеданных руд (*a*–*в*)
a – меломедистые; *б* – среднемедистые; *в* – высокомедистые

Рис. 21. Линии регрессии Cu-S для разных групп серноколчеданных руд

1-3 - руды: 1 - маломедистые; 2 - среднемедистые; 3 - высокомедистые; стрелки - вектора изменения двумерных средних

ная модель не может быть использована - значение эмпирического коэффициента корреляции между медью и серой в данном случае меньше критического.

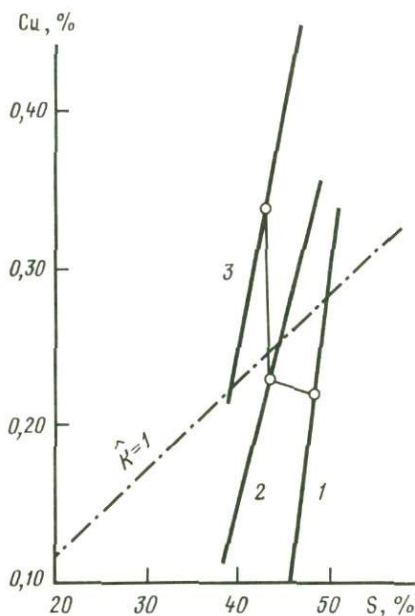
Общий вид поля регрессивной зависимости медь-сера можно видеть на рис. 20, в. В связи с невозможностью использования модели линейной зависимости между химическими элементами уравнений регрессии для меди по данной группе руд не приводятся.

Причины нарушения типичных для других групп серноколчеданных руд линейных зависимостей медь-сера в группе высокомедистых руд не вполне очевидны. Можно лишь предполагать, что в данном случае мы имеем дело с рудами, наиболее измененными последующими процессами, в том числе и полиметаллическим и медным рудообразованием.

Рассмотрим положение регрессионных линий зависимости содержаний меди от содержаний серы в общем поле регрессии медь-сера. На рис. 21 эти линии вынесены на общую диаграмму. Отчетливо видно, что частные линии регрессии медь-сера низкомедистых и среднемедистых серноколчеданных руд смещены одна относительно другой по оси абсцисс (уменьшение содержаний серы), а линия регрессии сера-медь высокомедистых серноколчеданных руд смещена относительно других линий регрессии вверх по оси ординат (обогащение медью), что и показано векторами изменения положения двумерных средних.

Изучение зависимости содержаний меди и серы в серноколчеданных рудах уральских колчеданных месторождений с помощью предлагаемой методики регрессионного анализа позволило сделать некоторые выводы относительно изменения состава серноколчеданных руд во времени. 1-я группа серноколчеданных руд (меломедистая) отражает, очевидно, состав первичных неизменных протопиритов. Две другие группы руд отражают разную степень изменения первичных руд в процессе последующих медного и полиметаллического этапов рудения.

На первом этапе изменения протопиритов, в стадию предрудного выщелачивания [12], происходит частичное выщелачивание рудного вещества (пирита), о чем может свидетельствовать снижение содержаний серы с 48,05 до 43,93% при неизменных содержаниях меди (0,22% - в протопиритах и 0,23% - в серноколчеданных рудах стадии предрудного выщелачивания).



При последующем наложении медного и полиметаллического орудения при малоизменяющихся содержаниях серы (43,93% — в стадию предрудного выщелачивания и 43,05% — в медно-полиметаллический этап) содержания металлов увеличиваются. Для меди рост содержаний происходит с 0,23% до 0,34%. При этом нарушается линейная зависимость содержаний серы и меди.

В ы в о д ы. Предлагаемая методика регрессивного анализа в приложении к исследованию взаимоотношений значений пары признаков в наборе геологических объектов позволяет выявить такие особенности поведения признаков (а также изучаемых объектов), которые не могут быть столь наглядно и эффективно установлены при применении методов математической статистики, традиционных для обработки геологической информации. Так, только с помощью методики построения прямых ортогональной регрессии и последующего анализа отклонений от них может быть выявлена неоднородность общей выборки относительно линии регрессии, что наглядно продемонстрировано на примере изучения динамики состава серноколчеданных руд. Столь же очевидна эффективность применения такой методики при разделении дифференцированных базальтоидных серий по близости характера изменения содержаний K_2O (по градиенту изменения) на группы, соответствующие петрологическим типам базальтоидного магматизма. Кроме того, эта методика позволила выявить и наглядно показать еще одно обстоятельство в пользу гипотезы о щелочно-базальтовых магмах как дериватах при дифференциации магм известково-щелочного типа. Не менее эффективно применение предлагаемой методики построения ортогональных регрессионных прямых при анализе совокупностей, объединяющих составляющие с различным характером изменения признаков в процессе дифференциации (см. пример анализа базальтоидной серии Фиджи).

В приведенном изложении не рассматривались примеры анализа совокупностей с обратной зависимостью признаков и в условиях их логнормального распределения. Можно указать, что реализация методики при этом не имеет принципиальных отличий от продемонстрированной на примерах в данной главе.

Г л а в а V

РЕГРЕССИОННЫЙ АНАЛИЗ ПРИ ИССЛЕДОВАНИИ ПОВЕДЕНИЯ НАБОРА ПРИЗНАКОВ В ОДНОМ ИЛИ НЕСКОЛЬКИХ ОБЪЕКТАХ

В гл. IV был рассмотрен регрессионный анализ при описании взаимоотношений пары признаков в совокупности геологических объектов. В данной главе обсуждается применение ортогонального регрессионного анализа при исследовании поведения набора признаков в одном или нескольких геологических объектах. В качестве примера осуществлен анализ зависимости содержаний элементов-спутников полиметаллической геохимической ассоциации колчеданных месторождений Мира от содержаний главного металла

Содержания (в г/т) металлов полиметаллической ассоциации
в колчеданных месторождениях Мира

Номер ц/л	Zn	Pb	Ba	Cd	Sb	In	Ge	Au	Ag
1	1300	100	300	10	50	1,5	1,0	0,50	6,3
2	2300	300	700	22	20	2,0	1,0	0,13	1,8
3	3100	500	600	15	10	2,0	0,5	0,25	2,1
4	3600	300	200	15	15	3,5	0,5	0,40	6,6
5	4000	800	2000	20	10	1,0	1,5	0,35	7,6
6	4500	200	400	20	15	2,0	0,7	0,27	3,7
7	4700	300	900	60	20	4,0	1,7	0,21	2,2
8	5000	2700	7000	50	30	1,6	0,7	0,56	10,8
9	10600	1100	2700	50	10	2,5	2,1	1,14	12,7
10	12000	200	2500	40	30	1,0	1,1	1,50	11,0
11	12000	1000	3100	120	75	1,5	3,0	1,33	21,6
12	13000	2400	7500	80	70	9,0	1,0	1,22	18,8
13	13200	1000	10500	80	110	1,0	2,0	0,36	9,5
14	13800	1300	3500	70	60	6,0	3,0	2,45	27,4
15	15700	1500	9000	128	350	5,0	1,9	0,34	3,0
16	17300	1700	13300	167	75	30,0	2,0	1,37	9,0
17	17900	4300	3600	177	200	18,0	3,0	1,57	26,4
18	18100	8100	8500	127	180	28,0	1,4	0,70	11,3
19	19300	400	117000	80	75	6,0	10,0	0,75	16,4
20	19400	7400	34100	141	95	10,0	1,0	0,48	12,2
21	19500	1500	8400	164	180	2,2	8,5	2,50	28,0
22	19900	700	14800	150	300	28,2	16,0	0,80	10,6
23	20400	900	48000	83	100	1,0	2,0	3,00	25,8
24	21500	900	46000	70	80	5,5	1,2	1,04	14,0
25	21600	9200	102000	80	400	33,0	20,0	8,60	76,5
26	25100	4400	11200	160	50	23,0	4,0	2,70	36,0
27	25100	6300	21200	163	100	9,0	1,4	1,43	10,3
28	25400	5800	12000	161	100	11,0	15,0	3,53	20,8
29	34500	900	40800	80	100	5,1	1,7	4,10	37,1
30	35700	4000	49000	242	120	8,0	9,3	2,58	47,5
31	36200	5000	2400	210	170	7,0	1,6	0,67	7,1
32	41600	8100	9000	320	150	26,0	2,0	1,15	23,7
33	45000	2700	85600	420	150	4,5	12,0	1,36	46,0
34	48100	5300	2800	325	250	4,0	8,8	1,90	27,5
35	55900	300	31400	380	200	17,0	14,0	3,30	33,8
36	61100	13100	71000	210	200	11,6	2,4	6,80	110,6
37	63900	6100	32000	400	95	12,0	4,0	4,40	108,0
38	64000	4600	52000	190	250	8,0	12,0	5,90	77,1
397	74300	14300	11500	490	100	13,0	25,0	2,25	34,2

Распределение содержаний металлов (в г/т) полиметаллической ассоциации
в колчеданных месторождениях

Статистические оценки	Символы	Цинк, г/т	Свинец, г/т	Барий, г/т
Число наблюдений	n	39	39	39
Среднее значение логарифмов	$\bar{\lg}x, \bar{\lg}y$	4,221	3,217	3,912
Стандартное отклонение	σ_{\lg}	0,431	0,580	0,731
Доверительный интервал среднего	λ	$\pm 0,138$	$\pm 0,186$	$\pm 0,234$
Показатель асимметрии распределения	$\Pi_{ас}$	+0,231	-0,026	+0,179
Показатель эксцесса распределения	$\Pi_{экс}$	-0,026	+0,333	+0,333
Толерантные пределы				
а) минимальные	T_{min}	3,359	2,057	2,450
б) максимальные	T_{max}	5,083	4,377	5,374

данной ассоциации — цинка, а также изучение зависимости содержаний редкоземельных металлов в магматических породах от значений показателя общей кислотности магм A_c .

Исследование геохимических спектров колчеданных месторождений [18] позволило выявить типовые для разных рудных формаций ассоциации химических элементов: серноколчеданную, медную и полиметаллическую. Объектом проведенного ниже изучения является последняя из перечисленных ассоциаций, главным металлом в которой выступает цинк. В качестве элементов-спутников цинка в данной ассоциации участвуют свинец, барий, кадмий, сурьма, серебро, золото, индий и германий. Между содержаниями цинка и содержаниями его спутников как представителями одной природной ассоциации должна существовать и действительно наблюдается положительная (прямая) корреляционная зависимость. Регрессионный анализ позволил не только наглядно представить характер этой зависимости, но и внести некоторые уточнения в наши представления о взаимоотношениях цинка и его элементов-спутников в рассматриваемом типе оруденения.

Регрессионный анализ при решении данной задачи осуществлялся на основе той же методики, что составляла содержание предшествующего изложения. Содержания всех металлов взяты в вычислениях в сопоставимых единицах — в граммах на тонну руды. Число наблюдений во всех процедурах соответствует числу месторождений мировой выборки, для которых подсчитаны средневзвешенные содержания металлов — 39 (исходные данные представлены в табл. 29). Как видно из параметров распределения содержаний (табл. 30), эмпирическое распределение содержаний всех элементов-спутников и цинка не противоречит логнормальной модели распределения, а выборочные совокупности для всех металлов в принятом методикой смысле однородны ($\Pi_{экс} < 0,500$).

При реализации процедуры регрессионного анализа в первую очередь было установлено, что во всех случаях в паре признаков цинк—элемент-спутник при построении линий регрессии содержания цинка следует ис-

Кадмий, г/т	Сурьма, г/т	Индий, г/т	Германий, г/т	Золото, г/т	Серебро, г/т
39	39	39	39	39	39
1,999	1,884	0,752	0,446	0,088	1,200
0,434	0,443	0,465	0,476	0,459	0,450
±0,139	±0,142	±0,149	±0,152	±0,147	±0,144
+0,077	+0,179	+0,026	-0,179	+0,026	+0,026
+0,128	+0,128	+0,282	+0,282	+0,333	+0,179
1,131	0,998	-0,178	-0,506	-0,830	0,300
2,886	2,770	1,682	1,398	1,006	2,100

пользовать как аргумент ($D_{Zn} < 1,000$). Вычисленные оценки коэффициента парной корреляции во всех случаях превышают минимально допустимое значение, т.е. для описания зависимости между содержаниями цинка и его элементов-спутников может быть использована модель прямой линейной регрессии. Как показывают результаты регрессионного анализа (табл. 31), углы наклона линий регрессии к оси абсцисс меняются для разных элементов-спутников от $45,4^\circ$ (кадмий) до $75,9^\circ$ (барий), ниже приводятся эмпирические уравнения регрессии для определения содержания элементов спутников полиметаллической ассоциации по содержаниям цинка в колчеданных месторождениях

$$\begin{aligned} \lg \text{Pb} &= 3,217 + 1,346 \cdot (\lg \text{Zn} - 4,221) \pm 0,886; \\ \lg \text{Ba} &= 3,912 + 1,696 \cdot (\lg \text{Zn} - 4,221) \pm 1,010; \\ \lg \text{Cd} &= 1,999 + 1,007 \cdot (\lg \text{Zn} - 4,221) \pm 0,342; \\ \lg \text{Sb} &= 1,884 + 1,028 \cdot (\lg \text{Zn} - 4,221) \pm 0,652; \\ \lg \text{In} &= 0,752 + 1,079 \cdot (\lg \text{Zn} - 4,221) \pm 0,856; \\ \lg \text{Ge} &= 0,446 + 1,104 \cdot (\lg \text{Zn} - 4,221) \pm 0,792; \\ \lg \text{Au} &= 0,088 + 1,065 \cdot (\lg \text{Zn} - 4,221) \pm 0,640; \\ \lg \text{Ag} &= 1,200 + 1,044 \cdot (\lg \text{Zn} - 4,221) \pm 0,608. \end{aligned}$$

Регрессионный анализ при изучении поведения совокупности признаков от одного, используемого как аргумент, сводится к графическому изображению на одной диаграмме всех линий регрессии, характеризующих указанные зависимости, с последующим их сопоставлением. Принцип графического изображения зависимости содержаний пары признаков и в данной задаче остается тем же, что и в рассмотренных выше случаях: шкалы значений признаков по координатным осям выбираются не произвольно, а в соответствии с приведенной выше методикой.

Параметры зависимости содержаний (в г/т) элементов-спутников
от содержаний (в г/т) цинка в колчеданных месторождениях

Статистические оценки	Символы	Свинец		Барий		Кадмий		Сурьма	
		Цинк	Свинец	Цинк	Барий	Цинк	Кадмий	Цинк	Сурьма
Число наблюдений	n	39		39		39		39	
Среднее значение логарифмов	$\lg x, \lg y$	4,221	3,217	4,221	3,912	4,221	1,999	4,221	1,884
Стандартное отклонение от среднего значения	σ	0,431	0,580	0,431	0,731	0,431	0,434	0,431	0,443
Коэффициент регрессии	\tilde{K}	0,743	1,346	0,590	1,696	0,993	1,007	0,973	1,028
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,329	0,443	0,298	0,505	0,170	0,171	0,317	0,326
Доверительный интервал для линии регрессии	$\tilde{\lambda}$	$\pm 0,105$	$\pm 0,142$	$\pm 0,095$	$\pm 0,162$	$\pm 0,054$	$\pm 0,055$	$\pm 0,102$	$\pm 0,104$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	+0,077		+0,077		+0,321		-0,128	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 0,658$	$\pm 0,886$	$\pm 0,506$	$\pm 1,010$	$\pm 0,340$	$\pm 0,342$	$\pm 0,634$	$\pm 0,652$
Логарифм коэффициента диспропорциональности	$\lg D$	-0,298	0,298	-0,600	0,600	-0,006	0,006	-0,024	0,024
Коэффициент диспропорциональности	D	0,504	1,986	0,251	3,981	0,986	1,014	0,946	1,057
Роль переменных		Аргумент	Функция	Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	63,3°		75,9°		45,4°		46,6°	
Критическое значение коэффициента корреляции	$R_{\text{крит.}}$	$\pm 0,320$		$\pm 0,320$		$\pm 0,320$		$\pm 0,320$	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	+0,618		+0,655		+0,803		+0,632	
Вид зависимости		Прямая		Прямая		Прямая		Прямая	

Т а б л и ц а 31 (окончание)

Статистические оценки	Символы	Индий		Германий		Золото	Серебро		
		Цинк	Индий	Цинк	Германий	Цинк	Золото	Цинк	Серебро
Число наблюдений	n	39		39		39	39		
Среднее значение логарифмов	$\lg x, \lg y$	4,221	0,752	4,221	0,446	4,221	0,088	4,221	1,200
Стандартное отклонение от среднего значения	σ	0,431	0,465	0,431	0,476	0,431	0,459	0,431	0,450
Коэффициент регрессии	\tilde{K}	0,927	1,079	0,905	1,104	0,939	1,065	0,958	1,044
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,397	0,428	0,358	0,396	0,300	0,320	0,291	0,304
Доверительный интервал для линии регрессии	$\tilde{\lambda}$	$\pm 0,127$	$\pm 0,137$	$\pm 0,115$	$\pm 0,127$	$\pm 0,096$	$\pm 0,102$	$\pm 0,093$	$\pm 0,097$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	+0,333		+0,128		+0,231		+0,333	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 0,794$	$\pm 0,856$	$\pm 0,716$	$\pm 0,792$	$\pm 0,600$	$\pm 0,640$	$\pm 0,582$	$\pm 0,608$
Логарифм коэффициента диспропорциональности	$\lg D$	-0,068	0,068	-0,090	0,090	-0,056	0,056	-0,038	0,038
Коэффициент диспропорциональности	D	0,855	1,169	0,813	1,230	0,879	1,137	0,916	1,091
Роль переменных		Аргумент	Функция	Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	49,5°		50,9°		48,7°		47,5°	
Критическое значение коэффициента корреляции	$R_{\text{крит.}}$	$\pm 0,320$		$\pm 0,320$		$\pm 0,320$		$\pm 0,320$	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	$\pm 0,540$		$\pm 0,584$		$\pm 0,651$		$\pm 0,662$	
Вид зависимости		Прямая		Прямая		Прямая		Прямая	

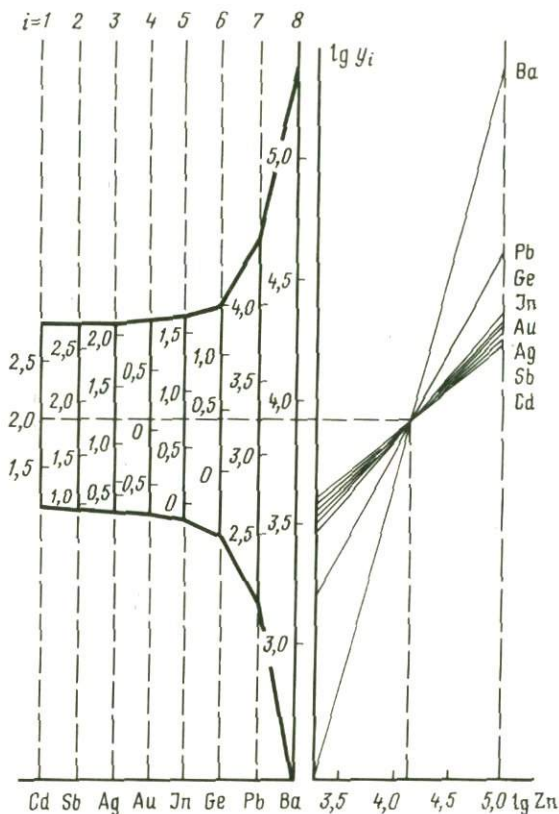


Рис. 22. Зависимость логарифмов содержаний элементов полиметаллической ассоциации колчеданных месторождений от логарифмов содержаний цинка

Первоначально и произвольно выбирается шкала для общего признака — признака-аргумента (в нашем случае для цинка). Шкалы значений для других признаков рассчитываются. Так как в данной задаче целью является сопоставление положения линий регрессии разных признаков по отношению к одному, то для каждого изучаемого признака-функции рассчитывается своя шкала значений по оси ординат на общей диаграмме. Подобный расчет необходим еще и потому, что в общем случае на ось ординат могут наноситься не только значения содержаний химических элементов, но и размеры рудных тел, температуры формирования минералов, расстояния от какого-либо реперного геологического объекта, значения геофизических параметров, плотность пространственного распределения месторождений полезных ископаемых и т.п.

Принцип расчета шкал для двух признаков был детально рассмотрен нами в главе III. Здесь же демонстрируется лишь принцип такого расчета для множества признаков по оси ординат (признаков-функций). Обратимся к рис. 22, где дано положение линий регрессии элементов-спутников

полиметаллической ассоциации колчеданных месторождений в общем поле зависимости от содержаний главного металла ассоциации – цинка.

По оси абсцисс произвольно взята шкала значений цинка (признака-аргумента). Затем в поле зависимости восстанавливаются три перпендикуляра: ограничительные линии минимальных и максимальных значений и линия средних значений цинка. Далее выбирается признак-функция с наибольшим рассчитанным углом наклона линии регрессии к оси абсцисс (в данном случае – бария) и под расчетным углом проводится искомая линия регрессии так, чтобы точки пересечения линии регрессии бария с ограничительной линией минимальных значений цинка была выше оси абсцисс. Точка пересечения линии регрессии бария с ограничительной линией минимальных значений цинка соответствует на оси ординат минимальным значениям бария, точка пересечения линии регрессии бария с ограничительной линией максимальных значений цинка соответствует на оси ординат максимальным значениям бария, тогда точка пересечения линии регрессии бария с линией среднего значения цинка будет соответствовать на оси ординат среднему значению бария. Имея на оси ординат отсчет минимальных, средних и максимальных значений бария, рассчитывают шкалу его значений.

Точка среднего значения на оси ординат одного признака-функции принимается, согласно методике, за точку средних значений всех остальных признаков-функций, т.е. все линии регрессии пересекаются в одной точке. Проведя линии регрессии остальных признаков-функций через общую точку средних значений под соответствующими рассчитанными углами и имея вследствие проделанной операции на оси ординат по три отсчета значений всех признаков-функций, рассчитывают для них требуемые шкалы значений. Для удобства рассчитанные шкалы выносят на вспомогательные линии левее оси ординат.

Проанализируем положение линий регрессии элементов-спутников полиметаллической ассоциации колчеданных месторождений на общей диаграмме. Обратимся еще раз к рис. 22: можно отметить, что все элементы-спутники цинка по градиенту возрастания их значений разделились на две группы – 1-я группа – кадмий, сурьма, серебро, золото, индий и германий; углы наклона линий регрессии этих металлов близки к 45° , т.е. изменение их содержаний происходит почти пропорционально изменению содержаний главного металла ассоциации – цинка и 2-я группа – свинец и барий; углы наклона линий регрессии этих металлов ($63,3^\circ$ и $75,9^\circ$ соответственно) свидетельствуют о том, что градиент изменения содержаний свинца и бария резко отличается от такового для цинка, т.е. пропорциональность изменения содержаний цинка, с одной стороны, и свинца и бария – с другой, резко нарушается.

Металлы первой группы (кадмий, сурьма, серебро, золото, индий и германий) в колчеданных месторождениях не образуют самостоятельных рудных тел, а находятся главным образом в рассеянном состоянии в минералах главного металла ассоциации – в сфалерите. Причем эти металлы по углам наклона линий регрессии к оси ординат (по градиенту изменения) образуют ряд, не противоречащий ряду геохимического сродства их с цинком.

Параметры зависимости содержаний (в г/т) редкоземельных металлов
в изверженных породах от показателя их общей кислотности A_c

Статистические оценки	Символы	Лантан	
		A_c	La
Число наблюдений	n	33	
Среднее значение	\bar{x}, \bar{y}	1,10	37,79
Стандартное отклонение от среднего значения	σ	0,30	16,26
Коэффициент пропорциональности	\tilde{K}	0,029	34,46
Коэффициент регрессии	\tilde{K}	0,019	53,67
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,135	7,24
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,046$	$\pm 2,47$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{\Pi}_{\text{экс}}$	-0,212	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 0,264$	$\pm 14,18$
Коэффициент диспропорциональности	D	0,642	1,558
Роль переменных		Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	57,3°	
Критическое значение коэффициента корреляции	$R_{\text{крит}}$	$\pm 0,347$	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	$\pm 0,778$	
Зависимость		Прямая	

Минералы свинца и бария могут как входить в цинкосодержащие руды, так и формировать собственные рудные тела, т.е. эти металлы обладают по сравнению с металлами первой группы большей геохимической автономностью, и вследствие этого степень изменения их концентраций в большей степени изменена, чем концентрация цинка и близких к нему редких и рассеянных элементов-спутников.

Таким образом, анализ зависимости набора признаков в одном геологическом объекте (выборка колчеданных месторождений рассматривается в данном случае как один геологический объект) по предлагаемой методике ортогонального регрессионного анализа позволил выделить группы признаков функций по градиенту их изменения по отношению к главному признаку — признаку-аргументу, а сами признаки-функции выстроить в ряд по степени их сроства с главным признаком, т.е. решить вполне конкретную геологическую задачу.

На основе применения анализа ортогональной регрессии для описания характера изменения набора признаков нами были сопоставлены результаты изучения и наглядного представления зависимости по предлагаемой методике и с помощью построения линий регрессии широко известным методом наименьших квадратов. Такое сравнение было проведено на примере изучения зависимости содержаний редкоземельных металлов в магматических породах от положения их в ряду дифференциации, отображаемого значением показателя общей кислотности магм.

Празеодим		Церий		Неодим	
A_c	Pr	A_c	Ce	A_c	Nd
25		33		28	
1,79	7,44	1,11	72,88	1,09	32,32
0,29	2,92	0,305	26,54	0,30	11,37
0,146	6,856	0,015	63,75	0,034	29,65
0,100	10,03	0,012	87,18	0,026	38,28
0,217	2,18	0,205	17,83	0,25	9,55
$\pm 0,085$	$\pm 0,85$	$\pm 0,070$	$\pm 6,08$	$\pm 0,092$	$\pm 3,54$
0,200		0,152		0,357	
$\pm 0,426$	$\pm 4,27$	$\pm 0,401$	$\pm 34,94$	$\pm 0,489$	$\pm 18,72$
0,684	1,463	0,731	1,368	0,775	1,291
Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
	55,6°		53,8°		52,2°
	$\pm 0,400$		$\pm 0,347$		$\pm 0,377$
	+0,626		+0,664		+0,580
Прямая		Прямая		Прямая	

По материалам, любезно представленным Л.С. Бородиным, были рассчитаны параметры ортогональной регрессии, характеризующие зависимость содержаний (кларков) девяти металлов редкоземельной группы — лантана, церия, празеодима, неодима, самария, гадолиния, диспрозия, иттербия и иттрия — от показателя общей кислотности магм (A_c). Параметры зависимости для всех пар признаков (содержания — A_c) представлены в табл. 32. Как следует из данной таблицы, зависимость содержаний всех изученных металлов и значений A_c может быть описана с помощью прямолинейной модели. С учетом известного фактора [6] о прямом соответствии значений A_c и содержаний кремнезема (SiO_2) можно утверждать, что для магматических объектов нормального ряда дифференциации содержания рассматриваемых редкоземельных металлов в магматических породах увеличиваются от основных разновидностей пород ($A_c = 0,6-0,7$) до кислых ($A_c = 1,5-1,7$). Несмотря на то что в анализируемые выборки объединены данные о содержаниях металлов в породах разных континентов, во всех случаях выборки признаны однородными относительно построенных линий регрессии. Это может служить основанием для вывода об устойчивости характера изменения содержаний редкоземельных металлов от степени дифференцированности магматических серий.

Для линий регрессии, характеризующих характер изменения содержаний каждого металла, были рассчитаны углы наклона (градиенты изменения содержаний) относительно оси абсцисс — A_c (x). Известно [2], что редко-

Т а б л и ц а 32 (окончание)

Статистические оценки	Символы	Самарий		Гадолиний	
		A_c	Sm	A_c	Gd
Число наблюдений	n	25		22	
Среднее значение	\bar{x}, \bar{y}	1,13	7,00	1,09	6,71
Стандартное отклонение от среднего значения	σ	0,268	2,10	0,31	2,09
Коэффициент пропорциональности	\hat{K}	0,161	6,195	0,162	6,188
Коэффициент регрессии	\tilde{K}	0,128	7,836	0,147	6,81
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,223	1,744	0,207	1,411
Доверительный интервал линии регрессии	$\tilde{\lambda}$	$\pm 0,092$	$\pm 0,719$	$\pm 0,087$	$\pm 0,590$
Показатель эксцесса распределения относительно линии регрессии	$\tilde{P}_{\text{экс}}$	0,280		0,546	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	$\pm 0,460$	$\pm 3,593$	$\pm 0,406$	$\pm 2,766$
Коэффициент диспропорциональности	D	0,791	1,265	0,909	1,100
Роль переменных		Аргумент	Функция	Аргумент	Функция
Угол наклона линии регрессии к оси абсцисс	α	51,7		47,6	
Критическое значение коэффициента корреляции	$R_{\text{крит}}$	$\pm 0,412$		$\pm 0,428$	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	+0,585		+0,663	
Зависимость		Прямая		Прямая	

земельные металлы по своим физико-химическим свойствам – электроотрицательности, кислотно-основным параметрам и т.п. – образуют ряд от лантана до иттрия (La, Ce, Pr, Nd, Sm, Gd, Dy, Yb, Y). На совмещенном графике зависимости (рис. 23), принцип построения которого рассматривался выше на примере полиметаллической ассоциации колчеданных месторождений, можно видеть, что линии регрессии, отражающие характер изменения содержаний металлов от значений A_c , по углу наклона к оси абсцисс образуют некоторый ряд. С учетом эмпирического характера изображенных линий этот ряд от линии регрессии с максимальным углом наклона к оси абсцисс (с наибольшим градиентом изменения содержаний) к линии с минимальным углом наклона имеет следующий порядок: La, Pr, Ce, Nd, Sm, Gd, Dy, Yb, Y. Нетрудно видеть, что этот ряд совпадает с приведенным по физико-химическим свойствам с точностью до "рокировки" местами церия с празеимом. Проверка гипотезы об одинаковом положении металлов в приведенных выше двух рядах (по характеру физико-химических свойств и по углу наклона эмпирических линий регрессии) с помощью рангового коэффициента корреляции Спирмена подтверждает сделанное предположение ($R_{x,y} = 0,983$ при допустимом значении $R_{\text{крит}} = 0,77$).

Диспрозий		Иттербий		Иттрий	
A_c	Dy	A_c	Yb	A_c	Y
16		29		33	
1,05	4,89	1,09	3,00	1,11	29,76
0,30	1,50	0,30	0,83	0,30	6,90
0,214	4,679	0,361	2,771	0,037	26,91
0,200	5,003	0,360	2,778	0,044	22,70
0,127	0,636	0,171	0,471	0,177	4,02
$\pm 0,064$	$\pm 0,318$	$\pm 0,062$	$\pm 0,171$	$\pm 0,060$	$\pm 1,37$
0,125		0,172		0,394	
$\pm 0,254$	$\pm 1,272$	$\pm 0,335$	$\pm 0,993$	$\pm 0,347$	$\pm 7,78$
0,935	1,069	0,995	1,005	1,186	0,843
Аргумент	Функция	Аргумент	Функция	Аргумент	Функция
46,9		45,1°		40,1°	
$\pm 0,500$		$\pm 0,370$		$\pm 0,347$	
+0,778		+0,715		+0,709	
Прямая		Прямая		Прямая	

Введем некоторые необходимые для геологических выводов предположения. Будем считать, что любезно предоставленные Л.С. Бородиным выборочные совокупности, составленные по анализам пород из разных континентов, отвечают следующему требованию: породы представляют магматические серии, процессы формирования которых принадлежат одному петрологическому типу. Косвенным подтверждением такого предположения является выявленная однородность наблюдений в выборках относительно линий регрессии. Предположим также, что совокупность физико-химических свойств каждого из изучаемых редкоземельных металлов прямо отражается на скорости возрастания его концентраций в продуктах каждой последующей стадии дифференциации магматического глубинного расплава. При этом следует допустить и линейность изменения этой скорости.

С учетом этих предположений ряд редкоземельных металлов, сформированный в соответствии с изменением угла наклона эмпирических линий регрессии на общей диаграмме, может рассматриваться как отражение последовательности тех же металлов, построенной на базе совокупности их физико-химических свойств. Основанием для подобного вывода, кроме приведенных предположений, которые в принципе не противоречат существ-

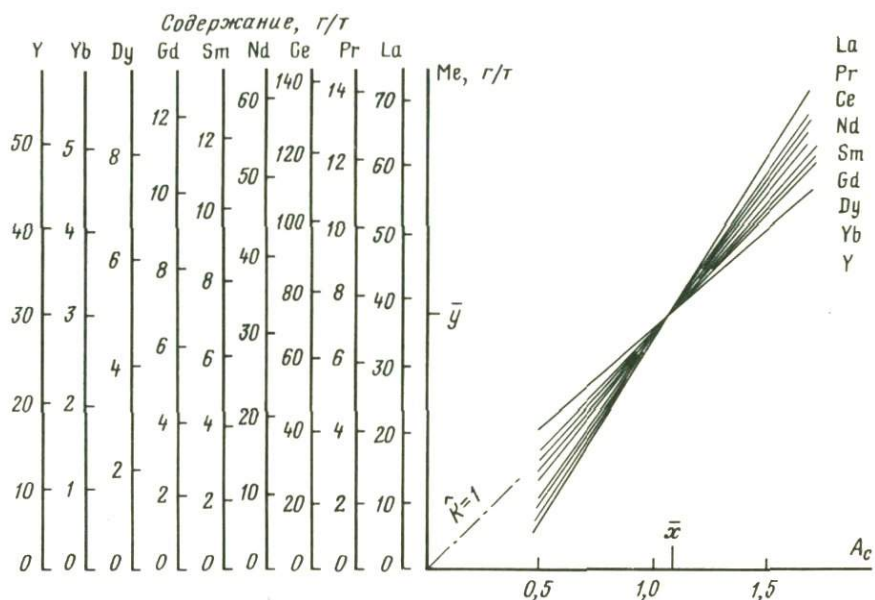


Рис. 23. Зависимость содержаний редкоземельных элементов от показателя общей кислотности пород (A_c), рассчитанная по программе "Lutar"

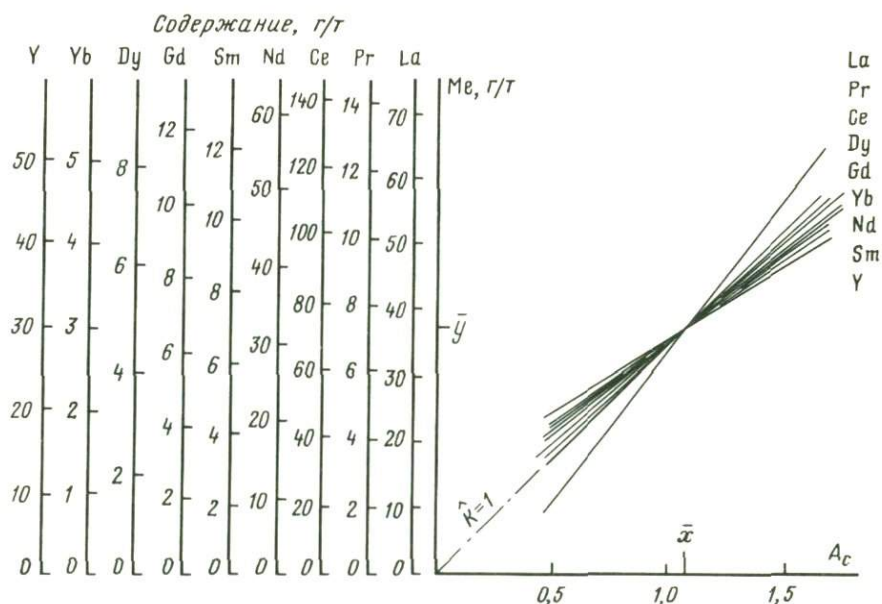


Рис. 24. Зависимость содержаний редкоземельных элементов от показателя общей кислотности пород (A_c), рассчитанная по программе "Aproks"

Ряды редкоземельных металлов

Номер п/п	По физико-химическим свойствам	По углу наклона линии ортогональной регрессии		По углу наклона линии регрессии метода наименьших квадратов		По величине оценки коэффициента парной корреляции ($C-A_c$)	
		металл	α , град	металл	α , град	металл	R_{C-A_c}
	1	2		3		4	
1	La	La	57,3	La	53,4	La	+0,864
2	Ge	Pr	55,6	Pr	44,5	Dy	+0,853
3	Pr	Ce	53,8	Ce	43,5	Yb	+0,809
4	Nd	Nd	52,2	Dy	42,4	Y	+0,806
5	Sm	Sm	51,7	Gd	39,1	Ce	+0,748
6	Gd	Gd	47,6	Yb	38,9	Pr	+0,737
7	Dy	Dy	46,9	Nd	38,8	Gd	+0,673
8	Yb	Yb	45,1	Sm	38,6	Sm	+0,630
9	Y	Y	40,1	V	34,3	Nd	+0,624

вующим представлениям о геохимии магматических процессов, является и практически совпадающее положение металлов в обеих рассматриваемых здесь упорядоченных на базе разных принципов последовательностях редкоземельных металлов.

Для сопоставления результатов применения различных методик, кроме приведенного выше ряда, сформированного по ортогональным линиям регрессии, на том же материале были выявлены аналогичные ряды металлов по убыванию угла наклона линий регрессии, вычисленных с помощью обычно применяемого метода наименьших квадратов и по убыванию значений оценок парного коэффициента корреляции содержаний редкоземельных металлов и значений A_c . Линии регрессии, построенные с помощью метода наименьших квадратов, изображены на рис. 24. Ряд редкоземельных металлов, упорядоченный по их физико-химическим свойствам, а также аналогичные ряды, построенные на основе обсуждаемых методов, приведены в табл. 33.

При анализе приведенных в таблице рядов металлов следует учитывать, что в качестве эталонного по рассмотренному выше геохимическому смыслу используется ряд редкоземельных металлов, упорядоченный по их физико-химическим свойствам (графа 2). Проверка гипотезы о совпадении положения металлов в эталонном ряду (графа 2) и в рядах, сформированных на основе статистического представления зависимости (графы 3–5), осуществлялась с помощью вычисления рангового коэффициента корреляции Спирмена, который в наибольшей степени отвечает смыслу этой задачи. Результаты вычисления оценки рангового коэффициента корреляции и проверки гипотезы о независимом положении металлов в рассматриваемой паре рядов сводятся к следующему: 1) ряд по физико-химическим свойствам и ряд по углу наклона линий ортогональной регрессии — $R_{1-2} = +0,983$; 2) ряд по физико-химическим свойствам и ряд по углу наклона линий регрессии методом наименьших квадратов — $R_{1-3} =$

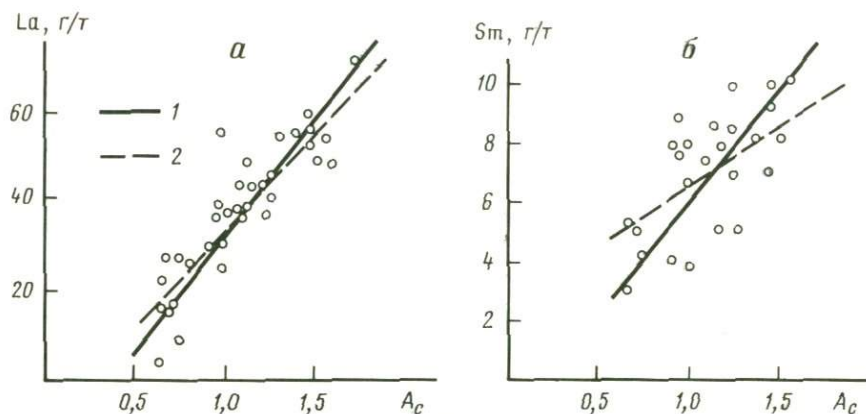


Рис. 25. Соотношение линии ортогональной регрессии (1) и линии регрессии метода наименьших квадратов (2) для зависимости содержаний лантана (а) и самария (б) от значений A_c

= + 0,717; 3) ряд по физико-химическим свойствам и ряд по величине оценки парной корреляции $R_{1-4} = -0,066$. Для всех случаев допустимое значение оценки коэффициента корреляции одинаково и равно $\pm 0,790$.

Из приведенных результатов следует, что близость к эталонному порядку в расположении редкоземельных металлов наблюдается только в ряду, сформированном по углу наклона линий регрессии, вычисленных по предлагаемой методике, т.е. по линиям ортогональной регрессии. Остановимся более подробно на особенностях применения результатов регрессионного анализа, осуществляемого, с одной стороны, предлагаемой методикой (реализована программой "Lutar") и методом наименьших квадратов – с другой (реализован программой "Arproks").

Рассмотрим эти особенности на примере зависимости от A_c содержаний двух металлов: лантана, для которого оценка коэффициента корреляции содержаний с A_c наибольшая и самария – с одной из наименьших оценок такого коэффициента. Точечные диаграммы и графическое изображение зависимости с помощью линий регрессии $C-A_c$ (по методу ортогональной регрессии и методу наименьших квадратов) для этих металлов приведены на рис. 25.

Для содержаний лантана (рис. 25,а), точки наблюдений для которого образуют на диаграмме четкий вытянутый ряд, положение эмпирических линий регрессии, вычисленных как по предлагаемой методике (сплошная линия), и методом наименьших квадратов (пунктирная линия), практически не отличается. В то же время значительно более существенное различие в положении подобных линий наблюдается на диаграмме: содержания самария – значения A_c (рис. 25,б). В случае с самарием ряд точек наблюдения имеет более округлую форму, вследствие чего линия регрессии по методу наименьших квадратов в большей степени не совпадает с линией регрессии по предлагаемой методике. Если линия ортогональной регрессии отражает только градиент изменения значений признаков, то линии регрессии по методу наименьших квадратов, как было показано в главе I, отражают и градиент изменения значений признаков, и степень разбросан-

ности точек наблюдения около средней линии зависимости, что приводит к отклонению этой линии в ту (y по x) или другую (x по y) сторону от ее среднего значения. А так как величина оценки коэффициента корреляции содержания A_c для разных металлов неодинакова, то и неравнозначное отклонение линии регрессии y по x от среднего положения приводит к нарушению ряда металлов по углу линий регрессии по методу наименьших квадратов.

Рассмотренные в данной главе примеры наглядно иллюстрируют применение предлагаемой методики при анализе поведения разных признаков в зависимости от изменения значений одного из них или от изменения значений любого смыслового параметра. Предложенный простой и надежный в условиях обработки метод построения эмпирических линий ортогональной регрессии, сопровождающийся методикой рационального выбора шкал и изображения результатов на единой диаграмме, совмещающей линии регрессии разных признаков в едином приведенном масштабе, обеспечивает возможность обоснования выводов в геохимических исследованиях самых разнообразных аспектов.

В последнем примере показано также, что положение линий ортогональной регрессии является более устойчивым по сравнению с положением линий регрессии наименьших квадратов, так как положение первых не зависит от степени разбросанности точек наблюдения в исследуемом рое.

ЗАКЛЮЧЕНИЕ

Ход геологического развития земной коры, формирование комплексов магматических, осадочных, метаморфических пород, а также образование разнообразных месторождений полезных ископаемых наряду с прочими факторами определяются временем формирования, пространственной локализацией, статикой (составом) и динамикой геологических процессов. Если время, место в пространстве и состав геологических образований фиксируются в большинстве случаев практически однозначно, то динамику этих процессов исследователи имеют возможность изучать лишь косвенно — по имевшим место в далеком геологическом прошлом изменениям различных признаков: состава, мощности, температуры, давления и т.п.

Представление о геологическом процессе (особенно эндогенном) как о реализации природной системы существенно повышает значимость изучения и анализа характера и изменчивости вещественных характеристик (признаков) в наблюдаемых продуктах деятельности данной системы. Причины, характер и комплекс условий проявления геологического процесса (системы) неизвестны исследователю, однако очевидно, что все эти ключевые факторы, идентифицирующие данный процесс, однозначно отображены в наблюдаемых объектах в совокупности и взаимоотношениях характеризующих их свойств (признаков). Естественно, что с учетом эффекта от мешающего влияния разнообразных побочных факторов большое значение в оценке характера геологического процесса имеет характер (в частности, скорость) изменения значений признаков (динамика) в пределах изучаемого объекта.

Нетрудно заключить, что относительная скорость (градиент) изменения значений одного из признаков по сравнению с динамикой изменения значений другого представляет собой количественную основу для косвенного изучения процессов формирования изучаемого объекта (объектов). Причем вполне очевидно, что данный подход к моделированию динамики свойств при реализации природной системы весьма продуктивен, особенно при необходимости типизации набора геологических объектов по характеру условий реализации процесса, отображенного в динамике взаимоотношений признаков.

Как показывает опыт исследований по выявлению взаимоотношений признаков, оценка градиентов изменения их значений с максимальной устойчивостью результата при минимальной степени искажения может быть обеспечена только при описании наблюдаемой зависимости с помощью метода ортогональной регрессии. Построение линий ортогональной регрессии позволяет с большой наглядностью и определенностью отразить характер градиента взаимного изменения значений признаков. Основанием для этого заключения является известное свойство ортогональной регрессии: на положение линий регрессии в двумерных диаграммах существен-

но меньшее влияние, чем при представлении зависимости широко применяемым методом наименьших квадратов, оказывают побочные, мешающие факторы, совокупность которых порождает так называемые случайные флуктуации.

Используя результаты значительного по объему опыта применения математических методов для описания зависимостей между значениями геологических признаков, а также учитывая существующие представления о процессах формирования объектов геологии, вполне логично принять, что, во-первых, в любой эндогенной системе присутствуют признаки (причем предположительно в значительном числе), изменение значений которых при переходе от состояния к состоянию системы осуществляется в пропорциональных количествах, что приводит к эффективному применению линейной модели зависимости, во-вторых, и при слабом нарушении пропорциональности возможно использование линейной модели, применение ортогональной линейной регрессии вполне оправдано с позиций одного из главных методологических принципов применения математических методов в геологии: при применимости нескольких математических моделей для описания реального объекта следует использовать наиболее простую из них.

Таким образом, можно принять, что зависимость значений геологических признаков в большинстве случаев может быть описана прямолинейной моделью, что существенно облегчает задачи исследователей. В связи с вышеизложенным в данной работе на основе существующих принципов исследования ортогональной регрессии авторами предложены отвечающие содержанию задач геологии и более простые, а значит и более удобные для практического применения методы решения задач регрессионного анализа.

1. Предложены уравнения регрессии для построения одной линии, отражающей градиент взаимного изменения значений обоих признаков.

2. Дана методика определения роли переменных в их взаимоотношениях (аргумент или функция) по градиентам взаимного изменения значений признаков.

3. Разработана процедура определения параметров поля регрессии.

4. Предложена методика расчета шкал значений признаков по координатным осям двумерных диаграмм.

5. Разработан метод исследования зависимости двух признаков во множестве геологических объектов с возможным выделением однородных групп (типизацией) объектов по градиентам изменения значений признаков или по отношению частных линий регрессии к обобщенной по всем объектам линии регрессии.

6. Предложена методика исследования зависимости набора признаков в одном или нескольких геологических объектах с установлением степени сродства признаков по градиенту изменения их значений.

Разработанная авторами и предлагаемая в работе методика регрессионного анализа для описания зависимости геологических признаков реализована в программе для ЭВМ, которая успешно используется в ИМГРЭ МГ СССР.

Ниже приводится алгоритм программы "Lutar", предназначенной для исследования линейной зависимости геологических признаков.

АЛГОРИТМ ПРОГРАММЫ "LUTAR"
(ИССЛЕДОВАНИЕ ЛИНЕЙНОЙ ЗАВИСИМОСТИ ГЕОЛОГИЧЕСКИХ ПРИЗНАКОВ)

A. Распределение значений признаков согласуется с нормальной моделью распределения величин.

1. Оценка среднего арифметического (математического ожидания)

$$\bar{x} = \frac{\sum_1^n x_i / n; \bar{y} = \frac{\sum_1^n y_i / n .$$

2. Оценка показателя асимметрии распределения значений признаков

$$П_{ас_x} = 1 - \frac{n(<\bar{x})}{0,5n}; \quad П_{ас_y} = 1 - \frac{n(<\bar{y})}{0,5n} .$$

3. Оценка стандартного отклонения от среднего значения

$$\sigma_x = \sqrt{\frac{\sum_1^n (\Delta x_i)^2}{(n-1)}}; \quad \sigma_y = \sqrt{\frac{\sum_1^n (\Delta y_i)^2}{(n-1)}} .$$

4. Оценка показателя эксцесса распределения значений признаков

$$П_{экс_x} = \frac{n(\bar{x} \pm 0,5\sigma_x)}{0,5n} - 1; \quad П_{экс_y} = \frac{n(\bar{y} \pm 0,5\sigma_y)}{0,5n} - 1 .$$

5. Оценка минимального значения толерантного предела значений признаков

$$T_{\min_x} = \bar{x} - t\sigma_x; \quad T_{\min_y} = \bar{y} - t\sigma_y .$$

6. Оценка максимального значения толерантного предела значений признаков

$$T_{\max_x} = \bar{x} + t\sigma_x; \quad T_{\max_y} = \bar{y} + t\sigma_y .$$

7. Оценка доверительного интервала среднего значения при уровне значимости $(1 - q)$:

$$\lambda_x = \pm \frac{t\sigma_x}{\sqrt{n}}; \quad \lambda_y = \pm \frac{t\sigma_y}{\sqrt{n}} .$$

8. Оценка коэффициента пропорциональности

$$\hat{K}_x = \bar{x} / \bar{y}; \quad \hat{K}_y = \bar{y} / \bar{x} .$$

9. Оценка коэффициента регрессии

$$\tilde{K}_x = \sigma_x / \sigma_y; \quad \tilde{K}_y = \sigma_y / \sigma_x .$$

10. Оценка коэффициента диспропорциональности изменения значений признаков

$$D_x = \tilde{K}_x / \hat{K}_x; \quad D_y = \tilde{K}_y / \hat{K}_y .$$

11. Оценка отклонений точек наблюдения от линии регрессии при $R_{x,y} > 0$:

$$\tilde{\Delta}x_{i_1} = \Delta x_i - \tilde{k}_x \Delta y_i; \quad \tilde{\Delta}y_{i_1} = \Delta y_i - \tilde{k}_y \Delta x_i .$$

12. Оценка отклонений точек наблюдения от линии регрессии при $R_{x,y} < 0$:

$$\tilde{\Delta}x_{i_{II}} = \Delta x_i + \tilde{k}_x \Delta y_i; \quad \tilde{\Delta}y_{i_{II}} = \Delta y_i + \tilde{k}_y \Delta x_i .$$

13. Оценка стандартного отклонения точек наблюдений от линии регрессии при $R_{x,y} > 0$:

$$\tilde{\sigma}_{x_1} = \sqrt{\frac{\sum_1^n (\tilde{\Delta}x_{i_1})^2}{n-1}}; \quad \tilde{\sigma}_{y_1} = \sqrt{\frac{\sum_1^n (\tilde{\Delta}y_{i_1})^2}{n-1}} .$$

14. Оценка стандартного отклонения точек наблюдений от линии регрессии при $R_{x,y} < 0$:

$$\tilde{\sigma}_{x_{II}} = \sqrt{\frac{\sum_1^n (\Delta \tilde{x}_{i_{II}})^2}{n-1}}; \quad \tilde{\sigma}_{y_{II}} = \sqrt{\frac{\sum_1^n (\Delta \tilde{y}_{i_{II}})^2}{n-1}}.$$

15. Доверительный интервал линии регрессии при уровне значимости $(1-q)$ (по меньшим $\tilde{\sigma}$):

$$\tilde{\lambda}_x = \pm \frac{t \tilde{\sigma}_x}{\sqrt{n}}; \quad \tilde{\lambda}_y = \pm \frac{t \sigma_y}{\sqrt{n}}.$$

16. Толерантные пределы отклонений точек наблюдений от линии регрессии при уровне значимости $(1-q)$ (по меньшим $\tilde{\sigma}$):

$$\tilde{T}_x = \pm t \tilde{\sigma}_x; \quad \tilde{T}_y = \pm t \tilde{\sigma}_y.$$

17. Оценка показателя эксцесса распределения точек наблюдения относительно линии регрессии (по меньшим $\tilde{\sigma}$):

$$\tilde{\Pi}_{\text{экс}} = \frac{n_{(\pm 0,5 \tilde{\sigma}_y)}}{0,5n} - 1.$$

18. Оценка коэффициента корреляции поля регрессии (по меньшим $\tilde{\sigma}$):

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_y}{\sigma_y} \cdot \frac{1}{1+(D_y)^2}.$$

19. Оценка допустимого (критического) значения коэффициента корреляции при уровне значимости $(1-q)$:

$$R_{(1-q)}(f) = t/\sqrt{n}.$$

Б. Распределение значений признаков согласуется с нормальным распределением логарифмов значений:

1) оценка среднего логарифмов значений признаков

$$\bar{\lg x} = \frac{\sum_1^n \lg x_i}{n}; \quad \bar{\lg y} = \frac{\sum_1^n \lg y_i}{n}.$$

2) оценка показателя асимметрии распределения логарифмов значений признаков

$$\Pi_{\text{ас}_{\lg x}} = 1 - \frac{n_{(< \bar{\lg x})}}{0,5n}; \quad \Pi_{\text{ас}_{\lg y}} = 1 - \frac{n_{(< \bar{\lg y})}}{0,5n}.$$

3. Оценка стандартного отклонения логарифмов значений от среднего

$$\sigma_{\lg x} = \sqrt{\frac{\sum_1^n (\Delta \lg x_i)^2}{n-1}}; \quad \sigma_{\lg y} = \sqrt{\frac{\sum_1^n (\Delta \lg y_i)^2}{n-1}}.$$

4. Оценка минимального толерантного предела логарифмов значений признаков:

$$\Pi_{\text{экс}_{\lg x}} = \frac{n_{(\bar{\lg x} \pm 0,5 \sigma_{\lg x})}}{0,5n} - 1; \quad \Pi_{\text{экс}_{\lg y}} = \frac{n_{(\bar{\lg y} \pm 0,5 \sigma_{\lg y})}}{0,5n} - 1.$$

5. Оценка минимального толерантного предела логарифмов значений признаков:

$$T_{\text{min}_{\lg x}} = \bar{\lg x} - t \sigma_{\lg x}; \quad T_{\text{min}_{\lg y}} = \bar{\lg y} - t \sigma_{\lg y}.$$

6. Оценка максимального толерантного предела логарифмов значений признаков

$$T_{\max_{\lg x}} = \bar{\lg x} + t\sigma_{\lg x}; \quad T_{\max_{\lg y}} = \bar{\lg y} + t\sigma_{\lg y}.$$

7. Оценка доверительного интервала среднего значения логарифмов значений признаков при уровне значимости $(1-q)$:

$$\lambda_{\lg x} = \pm \frac{t\sigma_{\lg x}}{\sqrt{n}}; \quad \lambda_{\lg y} = \pm \frac{t\sigma_{\lg y}}{\sqrt{n}}.$$

8. Оценка коэффициента регрессии логарифмов значений признаков:

$$\tilde{K}_{\lg x} = \sigma_{\lg x} / \sigma_{\lg y}; \quad \tilde{K}_{\lg y} = \sigma_{\lg y} / \sigma_{\lg x}.$$

9. Оценка коэффициента диспропорциональности изменения логарифмов значений признаков

$$\lg D_x = 2(\sigma_{\lg x} - \sigma_{\lg y}); \quad \lg D_y = 2(\sigma_{\lg y} - \sigma_{\lg x}).$$

10. Оценка отклонений точек наблюдения от линии регрессии при $R_{x,y} > 0$:

$$\tilde{\Delta} \lg x_{iI} = \Delta \lg x_i - \tilde{k}_{\lg x} \Delta \lg y_i; \quad \tilde{\Delta} \lg y_{iI} = \Delta \lg y_i - \tilde{k}_{\lg y} \Delta \lg x_i.$$

11. Оценка отклонений точек наблюдения от линии регрессии при $R_{x,y} < 0$:

$$\tilde{\Delta} \lg x_{iII} = \Delta \lg x_i + \tilde{k}_{\lg x} \Delta \lg y_i;$$

$$\tilde{\Delta} \lg y_{iII} = \Delta \lg y_i + \tilde{k}_{\lg y} \Delta \lg x_i.$$

12. Оценка стандартного отклонения точек наблюдения от линии регрессии при $R_{x,y} > 0$:

$$\tilde{\sigma}_{\lg x I} = \sqrt{\frac{\sum_1^n (\tilde{\Delta} \lg x_{iI})^2}{n-1}}; \quad \tilde{\sigma}_{\lg y I} = \sqrt{\frac{\sum_1^n (\tilde{\Delta} \lg y_{iI})^2}{n-1}}.$$

13. Оценка стандартного отклонения точек наблюдения от линии регрессии при $R_{x,y} < 0$:

$$\tilde{\sigma}_{\lg x II} = \sqrt{\frac{\sum_1^n (\tilde{\Delta} \lg x_{iII})^2}{n-1}}; \quad \tilde{\sigma}_{\lg y II} = \sqrt{\frac{\sum_1^n (\tilde{\Delta} \lg y_{iII})^2}{n-1}}.$$

14. Доверительный интервал линии регрессии при уровне значимости $(1-q)$ (по меньшим $\tilde{\sigma}_{\lg}$):

$$\tilde{\lambda}_{\lg x} = \frac{t \tilde{\sigma}_{\lg x}}{\sqrt{n}}; \quad \tilde{\lambda}_{\lg y} = \frac{t \tilde{\sigma}_{\lg y}}{\sqrt{n}}.$$

15. Толерантные пределы отклонений точек наблюдения от линии регрессии при уровне значимости $(1-q)$ (по меньшим $\tilde{\sigma}_{\lg}$):

$$\tilde{T}_{\lg x} = \pm t \tilde{\sigma}_{\lg x}; \quad \tilde{T}_{\lg y} = \pm t \tilde{\sigma}_{\lg y}.$$

16. Оценка показателя эксцесса распределения точек наблюдения относительно линии регрессии (по меньшим $\tilde{\sigma}_{\lg}$):

$$\tilde{\Pi}_{\text{экс}_{\lg}} = \frac{n(\pm 0,5 \tilde{\sigma}_{\lg y})}{0,5n} - 1.$$

17. Оценка коэффициента корреляции поля регрессии (по меньшим $\tilde{\sigma}_{\lg}$):

$$R_{x,y} = 1 - \frac{\tilde{\sigma}_{\lg y}}{\sigma_{\lg y}} \frac{1}{1 + (D_y)^2}.$$

18. Оценка допустимого значения коэффициента корреляции

$$R_{(1-q)}(f) = t/\sqrt{n}.$$

ЛИТЕРАТУРА

1. Айвазян С.А. Статистическое исследование зависимостей. М.: Металлургия, 1968.
2. Балашов Ю.А. Геохимия редкоземельных элементов. М.: Наука, 1976.
3. Большой Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1965.
4. Бондаренко В.Н. Статистическое решение некоторых задач геологии. М.: Наука, 1970.
5. Бондаренко В.Н., Коган Р.И., Чолакян П.Г. Методические рекомендации по первичной математической обработке данных при геохимических поисках месторождений. М.: ИМГРЭ, 1984.
6. Бородин Л.С. Геохимия главных серий изверженных пород. М.: Недра, 1981.
7. Бородин Л.С., Лутков Р.И. Особенности распределения и индикаторная роль K и R/b в дифференцированных базальтоидных сериях // Геохимия. 1982. № 9.
8. Ван дер Варден Б.Л. Математическая статистика. М.: Изд-во иностр. лит., 1960.
9. Верховская Л.А. Об интерпретации выборочных коэффициентов корреляции, вычисленных для неоднородных геохимических совокупностей // Математические методы исследования в геологии: Экспресс-информ. М.: ВИЭМС, 1973. Вып. 2.
10. Виноградов А.П. Средние содержания химических элементов в главных типах изверженных пород земной коры // Геохимия. 1962. № 7.
11. Гольдин С.В. О проверке однородности совокупностей геологических объектов // Математические методы при геологических исследованиях в Западной Сибири. Тюмень, 1968.
12. Жабин А.Г. Две фации пиритов в вулканогенном колчеданном месторождении Сибай на Южном Урале // Докл. АН СССР. 1974. Т. 218, № 5.
13. Жабин А.Г., Лутков Р.И. Геохимический критерий связи колчеданного оруденения Урала с базальтоидным вулканизмом // Магматизм, метаморфизм и оруденение в геологич. истории Урала.: Тез. докл. III Урал. петрогр. совещ. Свердловск, 1974.
14. Заварицкий А.Н. О метаморфизме колчеданных месторождений // Изв. АН СССР. Сер. геол. 1943. № 3.
15. Иванов С.Н. Новые данные о генезисе колчеданных месторождений Среднего Урала // Там же. № 1.
16. Лутков Р.И. Типы первичных геохимических ореолов Зырянского колчеданно-полиметаллического месторождения (Рудный Алтай) // Критерии интерпретации первичных геохимических ореолов. М.: ИМГРЭ, 1981.
17. Миллер Р.Л., Кан Дж. С. Статистический анализ в геологических исследованиях. М.: Мир, 1965.
18. Овчинников Л.Н., Лутков Р.И. Геохимические типы и зональность колчеданного оруденения Урала. М.: Наука, 1983.
19. Родионов Д.А. Функции распределения содержаний элементов и минералов в изверженных горных породах. М.: Наука, 1964.
20. Родионов Д.А. Статистические решения в геологии. М.: Недра, 1981.
21. Смирнов Н.В., Душин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1965.
22. Уилкс С. Математическая статистика. М.: Наука, 1967.
23. Imbric I. The place of biometric in taxonomy. H. Bull. Amer. Mus. Natur. Hist. 1956. Vol. 108, N 2.
24. Kermack K.A., Haldone I.B.S. Organic correlation and allometry // Biometrika, 1950. N 37.
25. Pearson K. On the lines and planes of closets in to system of points in space. // Phil. Mag. 1901. Vol. 6, N 2.

ПРИЛОЖЕНИЯ

Приложение 1

Допустимые значения критерия Стьюдента (*t*-критерий)
при различных уровнях значимости

Число степеней свободы, $(N - 2)$	Уровень значимости			Число степеней свободы $(N - 2)$	Уровень значимости		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,31	12,71	63,66	18	1,73	2,10	2,88
2	2,92	4,30	9,93	19	1,73	2,09	2,86
3	2,35	3,18	5,84	20	1,73	2,09	2,85
4	2,13	2,78	4,60	21	1,72	2,08	2,83
5	2,02	2,57	4,03	22	1,72	2,07	2,82
6	1,94	2,45	3,71	23	1,71	2,07	2,81
7	1,90	2,37	3,50	24	1,71	2,06	2,80
8	1,86	2,31	3,36	25	1,71	2,06	2,79
9	1,83	2,26	3,25	26	1,71	2,06	2,78
10	1,81	2,23	3,17	27	1,70	2,05	2,77
11	1,80	2,20	3,11	28	1,70	2,05	2,76
12	1,78	2,18	3,06	29	1,70	2,04	2,76
13	1,77	2,16	3,01	30	1,70	2,04	2,75
14	1,76	2,15	2,98	40	1,68	2,02	2,70
15	1,75	2,13	2,95	60	1,67	2,00	2,66
16	1,75	2,12	2,92	100	1,66	1,98	2,62
17	1,74	2,11	2,90		1,65	1,96	2,58

Приложение 2

Критические значения коэффициентов корреляции

Число степеней свободы $(N - 2)$	Вероятность ошибки первого рода (уровень значимости)			Число степеней свободы $(N - 2)$	Вероятность ошибки первого рода (уровень значимости)		
	0,10	0,05	0,01		0,10	0,05	0,01
1	0,988	0,997	1,000	17	0,389	0,456	0,575
2	0,900	0,950	0,990	18	0,378	0,444	0,561
3	0,805	0,878	0,959	19	0,369	0,433	0,549
4	0,729	0,811	0,917	20	0,360	0,423	0,537
5	0,669	0,754	0,874	25	0,323	0,381	0,487
6	0,621	0,707	0,834	30	0,296	0,349	0,449
7	0,582	0,666	0,798	35	0,275	0,325	0,418
8	0,549	0,632	0,765	40	0,257	0,304	0,393
9	0,521	0,602	0,735	45	0,243	0,287	0,372
10	0,497	0,576	0,708	50	0,231	0,273	0,354
11	0,476	0,553	0,684	60	0,211	0,250	0,325
12	0,457	0,532	0,661	70	0,195	0,232	0,302
13	0,441	0,514	0,641	80	0,183	0,217	0,283
14	0,426	0,497	0,623	90	0,173	0,205	0,267
15	0,412	0,482	0,606	100	0,164	0,195	0,254
16	0,400	0,468	0,590				

Приложение 3

Коэффициенты диспропорциональности и углы наклона линий регрессии

D (tg α)	α°	D (tg α)	α°	D (tg α)	α°
0,000	0,0	0,800	38,7	4,25	76,8
0,100	5,7	0,825	39,5	4,50	77,5
0,15	8,5	0,850	40,4	4,75	78,1
0,20	11,3	0,875	41,2	5,00	78,7
0,25	14,0	0,900	42,0	6,0	80,5
0,30	16,7	0,925	42,8	7,0	81,9
0,35	19,3	0,950	43,5	8,0	82,9
0,40	21,8	0,975	44,3	9,0	83,7
0,45	24,2	1,000	45,0	10,0	84,3
0,50	26,6	1,25	51,3	15,0	86,2
0,525	27,7	1,50	56,3	20,0	87,1
0,550	28,8	1,75	60,3	30,0	88,1
0,575	29,9	2,00	63,4	40,0	88,6
0,600	31,0	2,25	66,0	50,0	88,9
0,625	32,0	2,50	68,2	60,0	89,0
0,650	33,0	2,75	70,0	70,0	89,2
0,675	34,0	3,0	71,6	80,0	89,3
0,700	35,0	3,25	72,9	90,0	89,4
0,725	35,9	3,50	74,1	100,0	89,4
0,750	36,9	3,75	75,1	~	90
0,775	37,8	4,00	76,0		

Приложение 4. Макет

Таблица

Параметры распределения.

Статистические оценки	Символы	Признак 1	Признак 2
Число наблюдений	n	00	00
Среднее значение	\bar{x}, \bar{y}	0,00	0,00
Стандартное отклонение	σ	0,00	0,00
Доверительный интервал среднего	λ	$\pm 0,00$	$\pm 0,00$
Показатель асимметрии распределения	$P_{ас}$	$\pm 0,00$	$\pm 0,00$
Модель распределения		Нормальная (логнормальная)	Нормальная Логнормальная
Показатель эксцесса распределения	$P_{экс}$	$\pm 0,00$	$\pm 0,00$
Однородность выборки		Однородная (неоднородная)	Однородная Неоднородная
Толерантные пределы значений			
а) минимальные	T_{min}	0,00	0,00
б) максимальные	T_{max}	0,00	0,00

Параметры зависимости

Статистические оценки	Символы	Признак 1	Признак 2
Число наблюдений	n	00	
Среднее значение	\bar{x}, \bar{y}	0,00	0,00
Стандартное отклонение	σ	0,00	0,00
Коэффициент пропорциональности	\hat{K}	00,00	00,00
Коэффициент регрессии	\tilde{K}	00,00	00,00
Стандартное отклонение от линии регрессии	$\tilde{\sigma}$	0,00	0,00
Доверительный интервал линии регрессии	$\tilde{\lambda}$	±0,00	±0,00
Показатель эксцесса распределения относительно линии регрессии	$\tilde{\Pi}_{\text{экс}}$	±0,000	
Толерантные пределы отклонений от линии регрессии	\tilde{T}	±0,00	±0,00
Коэффициент диспропорциональности	D	0,00	0,00
Роль переменных		Аргумент (функция)	Функция Аргумент
Угол наклона линии регрессии к оси абсцисс	α	00,0°	
Критическое значение коэффициента корреляции	$R_{(1-q)}(f)$	±0,000	
Вычисленное значение коэффициента корреляции	$R_{x,y}$	±0,000	
Зависимость		Прямая (обратная)	

ОГЛАВЛЕНИЕ

Введение	3
Глава I. Особенности применения традиционных методов определения зависимости признаков	8
Глава II. Оценка параметров распределения переменных при исследовании их зависимости	16
Глава III. Регрессионный анализ при исследовании зависимости значений двух признаков.	38
Глава IV. Регрессионный анализ при исследовании поведения двух признаков в нескольких объектах.	61
Глава V. Регрессионный анализ при исследовании поведения набора признаков в одном или нескольких объектах.	96
Заключение.	112
Литература	117
Приложения	118

Научное издание

Лутков Рудольф Иванович
Бондаренко Валентин Николаевич

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ЗАВИСИМОСТИ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

*Утверждено к печати Институтом минералогии, геохимии
и кристаллохимии редких элементов*

Редактор *Т.Б. Гришина*. Художественный редактор *В.В. Алексеев*
Технический редактор *Г.П. Каренина*. Корректор *Н.И. Харламова*

Набор выполнен в издательстве на наборно-печатающих автоматах

ИБ № 39896

Подписано к печати 05.05.89. Т-10319. Формат 60 X 90 1/16. Бумага офсетная № 1

Печать офсетная. Гарнитура Пресс-Роман. Усл.печл.7,5. Усл.кр.-отт. 7,9

Уч.-издл. 8,4. Тираж 700 экз. Тип. зак. 1456. Цена 1 р. 70 к.

Ордена Трудового Красного Знамени издательство "Наука"
117864 ГСП-7, Москва В-485, Профсоюзная ул., д. 90

Ордена Трудового Красного Знамени 1-я типография издательства "Наука"
199034, Ленинград В-34, 9-я линия, 12

1 р. 70 к.

5154

