СОДЕРЖАНИЕ

[Введение 3](#_Toc64423348)

[1. Постановка задачи контрольной работы 4](#_Toc64423349)

[2. Многомерный корреляционный анализ данных в MS Excel 5](#_Toc64423350)

[3. Множественная регрессия в MS Excel 6](#_Toc64423351)

[Заключение 11](#_Toc64423352)

[Список использованных источников 12](#_Toc64423353)

**Введение**

Любое геологическое явление может быть охарактеризовано множеством признаков, поддающихся наблюдению и измерению. Геологические объекты должны рассматриваться как системы, зависящие от большого числа факторов и требующие для своего описания многомерного признакового пространства. Так, например, магматические породы сходного минерального и химического составов могут обладать некоторыми закономерностями, определяющими их специфическую рудоносность. Эти закономерности не поддаются выявлению с первого взгляда, однако они могут быть установлены путем статистической обработки результатов химических анализов пород. При решении таких задач необходимо совместное рассмотрение комплекса изучаемых признаков, то есть создание многомерной статистической модели.

В качестве математической модели значений комплекса признаков рассматривается многомерная случайная величина, которая часто называется случайным вектором. Многомерные модели подразумевают вероятность нормального статистического распределения рассматриваемых случайных величин или хотя бы возможности их нормализации. Однако статистические критерии для большинства процедур многомерного анализа разработаны при очень сильных ограничениях или основываются на логических соображениях.

Вследствие сложных стохастических взаимосвязей между изучаемыми признаками (переменными) часто не удается принять правильное решение относительно каждой из них. В таких случаях очень эффективно всестороннее исследование системы с выделением наиболее важных факторов, объединяющих влияние нескольких переменных.

Многомерные методы статистических исследований сложны. В большинстве многомерных геологических задач приходится иметь дело со сложными сочетаниями действующих факторов, которые не удается выделить в чистом виде и изучить изолированно. Но многомерные методы являются достаточными средствами геологических исследований, поскольку они позволяют геологу одновременно работать с большим числом переменных, чем он может осознать сам. Совместное изучение комплексов взаимосвязанных переменных (признаков) способствует выявлению дополнительной существенной информации об изменчивости свойств изучаемых объектов и обеспечивает возможность прогнозирования их неизвестных свойств.

1. **Постановка задачи контрольной работы**

В рудах полиметаллического месторождения кроме основных полезных компонентов – цинка, свинца и меди, содержатся попутные полезные компоненты – золото, серебро, кадмий, сурьма, барий, извлекаемые из руд в процессе переработки.

По условию задачи попутным элементов является Au (золото).

Требуется выполнить решение по следующим пунктам в программе MS Excel:

1. проверить гипотезу о наличии корреляционной связи между основными и одним из попутных компонентов;

1. получить корреляционную матрицу. Выделить значимые коэффициенты корреляции;
2. определить, с каким из основных компонентов наиболее тесно связан попутный компонент;
3. используя результаты корреляционного анализа при наличии корреляционных связей между основными и попутным компонентом, провести регрессионный анализ с учетом связей между основными и попутным компонентами. Рассчитать и записать уравнение регрессии для оценки содержания попутного компонента по основным компонентам.

**2. Многомерный корреляционный анализ данных в MS Excel**

Многомерный корреляционный анализ применяется для выявления зависимостей между обнаруженными значениями различных геологических характеристик и разделения множества признаков по характеру их внутренних связей.

**База данных для исследований**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № п/п | | Cu, % | | Pb, % | | Zn, % | | Au, г/т | |
| 1 | | 0,26 | | 1,73 | | 8,67 | | 0,2 | |
| 2 | | 0,2 | | 1,66 | | 4,47 | | 0,1 | |
| 3 | | 1,26 | | 3,29 | | 2,02 | | 0,6 | |
| 4 | | 0,34 | | 3,08 | | 8,46 | | 0,4 | |
| 5 | | 0,06 | | 0,21 | | 0,42 | | 0,2 | |
| 6 | | 0,11 | | 1,5 | | 3,2 | | 0,4 | |
| 7 | | 0,14 | | 1,6 | | 3,49 | | 0,1 | |
| 8 | | 0,09 | | 0,65 | | 1,7 | | 0,2 | |
| 9 | | 0,26 | | 2,05 | | 3,82 | | 0,2 | |
| 10 | | 0,29 | | 2,05 | | 4,66 | | 0,1 | |
| 11 | | 0,12 | | 1,43 | | 3,3 | | 0,1 | |
| 12 | | 0,02 | | 0,55 | | 1,85 | | 0,1 | |
| 13 | | 0,12 | | 0,25 | | 2,6 | | 0,1 | |
| 14 | | 0,38 | | 0,08 | | 5,53 | | 0,4 | |
| 15 | | 0,3 | | 0,14 | | 8,41 | | 0,4 | |
| 16 | | 0,02 | | 0,46 | | 1,76 | | 0,2 | |
| 17 | | 0,34 | | 3,08 | | 8,46 | | 0,4 | |
| 18 | | 1,26 | | 3,29 | | 22,82 | | 0,6 | |
| 19 | | 2,22 | | 0,66 | | 15,88 | | 0,1 | |
| 20 | | 0,75 | | 0,78 | | 4,2 | | 0,1 | |
| 21 | | 4,64 | | 0,37 | | 13,48 | | 0,4 | |
| 22 | | 3,64 | | 0,85 | | 35,97 | | 0,8 | |
| 23 | | 0,95 | | 0,56 | | 8,02 | | 0,2 | |
| 24 | | 0,09 | | 0,66 | | 1,47 | | 0,1 | |
| 25 | | 0,02 | | 0,46 | | 0,85 | | 0,1 | |
| № п/п | | Cu, % | | Pb, % | | Zn, % | | Au, г/т | |
| 26 | | 0,02 | | 0,39 | | 1,18 | | 0,1 | |
| 27 | | 0,15 | | 0,08 | | 2,9 | | 0,1 | |
| 28 | | 0,25 | | 0,06 | | 2,9 | | 0,1 | |
| 29 | | 1,17 | | 0,12 | | 9,25 | | 0,1 | |
| 30 | | 0,06 | | 0,06 | | 1 | | 0,1 | |
| 31 | | 0,05 | | 0,02 | | 1,58 | | 0,1 | |
| 32 | | 0,23 | | 0,09 | | 3,12 | | 0,1 | |
| 33 | | 0,09 | | 0,05 | | 0,63 | | 0,1 | |
| 34 | | 0,15 | | 0,12 | | 0,9 | | 0,1 | |
| 35 | | 0,06 | | 0,75 | | 1,71 | | 0,1 | |
| 36 | | 0,1 | | 0,1 | | 3,2 | | 0,1 | |
| 37 | | 0,44 | | 2,32 | | 8,2 | | 0,1 | |
| 38 | | 0,08 | | 0,49 | | 1,05 | | 0,1 | |
| 39 | | 0,02 | | 0,22 | | 0,65 | | 0,4 | |
| 40 | | 0,02 | | 0,46 | | 1,3 | | 0,1 | |
| 41 | | 0,02 | | 0,47 | | 0,94 | | 0,1 | |
| 42 | | 1,06 | | 5,61 | | 29,3 | | 0,8 | |
| 43 | | 0,58 | | 4,51 | | 18,28 | | 0,4 | |
| 44 | | 0,54 | | 3,41 | | 6,15 | | 0,4 | |
| 45 | | 0,11 | | 0,83 | | 1,92 | | 0,1 | |
| 46 | | 0,08 | | 3,21 | | 7,44 | | 0,4 | |
| 47 | | 0,52 | | 1,69 | | 3,3 | | 0,2 | |
| 48 | | 0,3 | | 7,02 | | 24,37 | | 0,2 | |
| 49 | | 0,26 | | 1,69 | | 2,6 | | 0,4 | |
| 50 | | 0,37 | | 5,61 | | 11,34 | | 0,4 | |

**Таблица зависимости элементов от попутного Au**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cu** | **Pb** | **Zn** |
| **Зависимость** | 0,5 | 0,5 | 0,7 |

При оценке силы связи коэффициентов корреляции используется шкала Чеддока:



1. **Множественная регрессия в MS Excel**

Множественная регрессия **—** линейная регрессия с двумя или более объясняющими переменными. Например, переменная у может быть выражена через переменные хi, при i=1, 2, ..., N. Тогда уравнение регрессии выглядит следующим образом: у=α+β1x1i+β2x2i+...+βNxNi+εi.

Модели множественной регрессии используются для предсказаний значений зависимой переменной (например, содержаний ценного элемента, объемной массы руды и глубины формирования минерала и др.) по набору независимых переменных (например, содержаний породообразующих элементов, объемных масс тяжелых минералов в рудах, содержаний элементов-индикаторов в минералах и др.).

Корреляционная зависимость по всему объему данных.

Проверяем зависимость всех компонентов в массиве друг к другу.

**Корреляционная матрица**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Cu** | **Pb** | **Zn** | **Au** |
| **Cu** | 1,0 |  |  |  |
| **Pb** | 0,0 | 1,0 |  |  |
| **Zn** | 0,6 | 0,6 | 1,0 |  |
| **Au** | 0,5 | 0,5 | 0,7 | 1,0 |

*Интерпретация результатов:*

*Коэффициент парной корреляции между Zn и Au, равный 0,7, показывает, что можно говорить о наличии высокой корреляционной связи. По остальным компонентам наблюдается слабая связь, что означает необходимость рассмотреть модели для всех компонентов. И на основании коэффициентов модели, выбрать наиболее достоверный.*

**Дисперсионный анализ**

Дисперсионный анализ оценивает достоверность полученной модели

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 3 | 0,867755192 | 0,289251731 | 15,70082145 | 3,61618E-07 |
| Остаток | 46 | 0,847444808 | 0,018422713 |  |  |
| Итого | 49 | 1,7152 |  |  |  |

По уровню значимости критерия Фишера (строка Регрессия, столбец

*Значимость F* 3,61618E-07, то есть p << 0,05 – заданный уровень значимости, гипотеза о незначимости регрессии отвергается и считается, что регрессия значима).

**Построение регрессионной статистики с помощью анализа данных по всем компонентам**

|  |  |
| --- | --- |
| *Регрессионная статистика* | |
| Множественный R | 0,711281031 |
| R-квадрат | 0,505920705 |
| Нормированный R-квадрат | 0,473698142 |
| Стандартная ошибка | 0,135730296 |
| Наблюдения | 50 |

\

По уровню значимости критерия Фишера (строка Регрессия, столбец

*Значимость F* 7,74667E-08, то есть p << 0,05 – заданный уровень значимости, гипотеза о незначимости регрессии отвергается и считается, что регрессия значима).

**Определение значения коэффициентов модели по компонентам**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |
| Y-пересечение | 0,1078 | 0,0274 | 3,9359 | 0,0003 |
| Cu, % | 0,0533 | 0,0329 | 1,6212 | 0,1118 |
| Pb, % | 0,0339 | 0,0165 | 2,0558 | 0,0455 |
| Zn, % | 0,0084 | 0,0045 | 1,8685 | 0,0681 |

Они определяются из таблицы в столбце **Коэффициенты** – в строке **Yпересечение** приводится свободный член 0,1078; в строках соответствующих переменных приводятся значения коэффициентов при этих переменных Cu= 0,0533, Pb=0,0339, Zn=0,0084 .

Проверим значимость коэффициентов регрессии, сравнивая попарно значения столбцов **Коэффициенты** и **Стандартная ошибка**. Можно заметить, что абсолютные значения коэффициентов больше, чем их стандартная ошибка. Эти коэффициенты являются значимыми.

Докажем значимость коэффициентов еще и по значениям показателя **Р-значение**, которые меньше заданного уровня значимости α=0,05. Коэффициент для Pb = 0,0455. Одно **Р-значение**, элементов Cu, и Zn больше заданного уровня значимости α=0,05. (это означает, что в уравнении регрессии оно будет нулевым)

Так как модель множественной регрессии имеет заметную детерминацию, выше отдельных моделей остановимся на ней.

Составим уравнение регрессии содержания попутного компонента по основным компонентам, используя полученные коэффициенты модели:

***Au (Cu, Pb, Zn)=* *0\*Cu + 0.0339\*Pb + 0\*Zn + 0,1078***

Статистическая модель построена

**Построение регрессионной статистики с помощью анализа данных по компоненту Pb**

|  |  |
| --- | --- |
| *Регрессионная статистика* | |
| Множественный R | 0,508917432 |
| R-квадрат | 0,258996953 |
| Нормированный R-квадрат | 0,243559389 |
| Стандартная ошибка | 0,162722183 |
| Наблюдения | 50 |

Степень описания моделью процесса – R-квадрат (вторая строка сверху в таблице Регрессионная статистика, в примере R-квадрат = 0,26*, что является недостоверным показателем модели).*

**Построение регрессионной статистики с помощью анализа данных по компоненту Сu**

|  |  |
| --- | --- |
| *Регрессионная статистика* | |
| Множественный R | 0,480781844 |
| R-квадрат | 0,231151182 |
| Нормированный R-квадрат | 0,215133498 |
| Стандартная ошибка | 0,165751414 |
| Наблюдения | 50 |

Степень описания моделью процесса – R-квадрат (вторая строка сверху в таблице Регрессионная статистика, в примере R-квадрат = 0,23, *что является*

**Построение регрессионной статистики с помощью анализа данных по компоненту Zn**

|  |  |
| --- | --- |
| *Регрессионная статистика* | |
| Множественный R | 0,674710606 |
| R-квадрат | 0,455234402 |
| Нормированный R-квадрат | 0,443885119 |
| Стандартная ошибка | 0,13952165 |
| Наблюдения | 50 |

Степень описания моделью процесса – R-квадрат (вторая строка сверху в таблице Регрессионная статистика, в примере R-квадрат = 0,46, *что является недостоверным показателем модели*).

**Заключение**

Метод построения математической модели может использоваться при поисках месторождений, а также других различных сравнений и зависимостей. Данный метод анализа корреляционной зависимости показывает, что в лабораторном исследовании проб, если есть зависимость одного элемента от другого, то можно предположить увеличение или уменьшения концентрации основного элемента полезного ископаемого от попутного.

Стоит обратить внимание на попутный минерал и при увеличении его концентрации можно спрогнозировать направление коренного месторождения содержащие основной минерал от которого зависит попутный. Коренное месторождение — это источник россыпей, в котором концентрация основного элемента полезного ископаемого высока по площади распространения.

Использование математической модели также имеет некую долю погрешности. Один из вариантов уменьшения погрешности – это зависимость числа проб при анализе, чем больше проб используют в статистике, тем меньше погрешность. Существуют также влияние внешних факторов, которые нельзя описать математическим путем. Внешне мы можем указывать на данный объект, что он имеет зависимость, увеличение или уменьшение, но произвести расчетный анализ очень сложно.