

地质数据统计分析

下 册

〔美〕 小乔治·S·科克 著
理查德·F·林克

科 学 出 版 社

内 容 简 介

美国小乔治·S·科克与理查德·F·林克合著的《地质数据统计分析》分上、下两册出版。上册包括一至三编共八章。本书为下册，包括第四至第六编，共九章。第四编讲多元统计方法：多重回归，广义方差分析，判别式分析，因子分析，并讨论了方向数据的分析，比与定和变量的问题；第五编是专门讲应用部分，特别讲到运筹学的基本思想与线性规划及其在地质数据分析中的应用。此外，还在这一部分比较详细地阐述了对数正态分布与金矿数据的关系；第六编只有一章，即第十七章，概述电子计算机与地质学。

可供已具有基本统计知识或读过本书上册的地质工作者参考，也可供对地质数据的统计分析感兴趣的其他领域的工作者参考。

George S. Koch, Jr. Richard F. Link

STATISTICAL ANALYSIS

OF GEOLOGICAL DATA

Vol. 2

John Wiley & Sons, Inc 1971

地质数据统计分析

下 册

〔美〕小乔治·S·科克 著
理查德·F·林克

刘 绂 堂 译

赵 鹏 大 校

·

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

·

1978年8月第一版 开本：787×1092 1/32

1980年6月第二次印刷 印张：13 7/8

印数：9,931—12,500 字数：317,000

统一书号：13031·788

本社书号：1132·13—14

定 价：1.85 元

作者序

本书是《地质数据统计分析》一书的下册¹⁾。目的是介绍比上册中更深一些的地质数据统计分析方法。当本书单独存在时，它又是前一册的逻辑继续。

本册是为具有相当于一年级大学课程所教的或通过上册的学习所获得的基本统计学知识的人写的。我们假定读者了解单变量观测值分析的统计方法，统计样本与总体的关系，及地质取样的基本原则。

我们重申上册作者序里如下一段话：我们的目的是，说明对地质数据分析有效的统计程序，并讨论取得有分析价值的可靠数据的方法。我们是为这样的人写的，他占有数据，需要从中作出结论，或是有一个问题，其解决需要取得数据和解释数据。我们侧重基本的统计方法，并强调其经过仔细思考的应用将产生有效的结果。因为地质学是一门内容复杂的综合性的科学，我们有目的地把某些必然牵涉到的地质论证放在本书中。可是，数学却比较简单，了解它们仅需要初等代数与几何，及矩阵代数的简单运算。

本册是为有些地质训练、也许有专业经验的读者写的。虽然大多数读者可能是地质工作者，本书也适于矿山及石油工程技术人员、地球化学工作者及地球物理工作者、矿物经济工作者、及其他人阅读。

虽然本册中大多数统计分析是真实数据的分析，有些却

1) 本书上册于1978年由科学出版社出版，有关翻译本书的介绍和说明，请参阅上册“译者的话”，这里就不加赘述——译者

是为说明专门统计方法的数值性质所选的虚构数据的分析。真实数据来自地质学中各个方面,特别是经济地质学、构造地质学及岩石学方面的数据。

本册包括全书的第四、五及六编。第四编说明多变量地质观测值的几种统计分析方法,并应用之于解决各种地质问题。第五编评述地质应用中某些问题,包括运筹学在地质及矿业中的几种应用及地质取样的某些专门方法。由于本册中的统计方法一般依靠电子计算机来实际执行,第六编是概述电子计算机在地质数据分析中的作用。

1970 年 6 月

目 录

作者序	i
-----------	---

第四编 多元统计方法

第九章 地质趋势	2
第一节 引言	2
第二节 一维数据的线性回归分析	8
第三节 相关系数	17
第四节 二维数据的线性与二次回归分析	20
第五节 二维数据回归分析举例	27
第六节 三维数据的线性与二次回归分析	40
第七节 三维数据回归分析举例	46
第八节 拟合趋势面的剩余	48
第九节 其他类型的图件分析	59
第十节 趋势分析的其他用途	66
第十一节 趋势分析的部署	74
参考文献	82
第十章 多元数据分析	85
第一节 引言	85
第二节 广义方差	92
第三节 多重回归	94
第四节 广义方差分析	106
第五节 判别式分析	111
第六节 因子分析	131
第七节 方向数据的分析	143
参考文献	164

第十一章	比与定和变量	166
第一节	引言	166
第二节	比	167
第三节	比的统计分析	174
第四节	定和变量	182
第五节	和为一常数的三变量数据的解释	186
	参考文献	200

第五编 应用地质学中的某些问题

第十二章	勘探	202
第一节	地区的地质分类	203
第二节	目标分布的模型	214
第三节	勘探网	222
第四节	响应面	229
第五节	搜索理论	234
	参考文献	239
第十三章	矿床的圈定	241
第一节	矿床圈定的某些条件	244
第二节	品位的计算	247
第三节	矿石量的计算	260
第四节	矿石估计的实例	267
第五节	矿石的品位与数量的关系	271
第六节	矿石储量的分类	277
	参考文献	280
第十四章	运筹学	281
第一节	引言	281
第二节	某些运筹学程序	286
第三节	找矿前阶段	304
第四节	勘探阶段	309
第五节	计算阶段	314

第六节	采掘阶段	319
第七节	制造	324
第八节	大面积调查	327
	参考文献	333
第十五章	专门地质取样	336
第一节	破碎岩石取样	337
第二节	砂矿取样	342
第三节	化探取样	351
第四节	对评价几种取样方法的研究	359
	参考文献	362
第十六章	金矿数据与对数正态分布	364
第一节	非洲南部西蒂迪普金矿数据实例	366
第二节	南达科他州里德地区霍姆斯特克金矿数据实例	373
第三节	其他金矿数据实例	380
第四节	金矿观测值与对数正态模型的偏离	382
第五节	小结	389
	参考文献	390

第六编 电子计算机与地质学

第十七章	电子计算机与地质学	392
第一节	计算机的应用	392
第二节	计算机程序设计	396
第三节	数据的获得、存储与恢复	402
第四节	编辑数据	406
第五节	数值方法	410
第六节	计算结果的显示	412
第七节	计算的部署	414
第八节	数字计算机的工艺	418
	参考文献	422

附录	423
----------	-----

部分符号表	426
中英人名对照表	428
中英地名、矿名及其他译名对照表	432

第四编 多元统计方法

第九至十一章有选择地介绍多元地质数据的统计分析方法。多元数据与一个观测值中仅有一个变量的单元数据不同，它们是一个观测值中有一个以上变量的数据。例如一种以上成分的岩石分析、一个化石标本的多种尺寸和形状的测定值、一种砂样上的多种沉积岩石的测定值、及化探中的多元素分析等。

第九章是讲地质趋势分析的，包括二维或三维观测值的数学模型表示，趋势面剩余值的解释及置信带与置信体的计算。第十章讲多元观测值的解释，所用的方法包括多重回归、广义方差分析、广义判别式分析与判别式函数、及因子分析。也考虑了一种特殊的多元观测值的，即方向数据的分析。第十一章讨论在多元分析中相互联系的三个问题：比，定和变量及三角形图的解释。

第九章 地质趋势

第一节 引言

地质趋势是一个地质变量或诸地质变量,例如花岗岩的长石含量或砂岩的孔隙度之类的值的系统变化,这种变化通常都是在一个地理方向上的。这个术语是由格兰特(1957)引入地质学的。克伦宾与惠顿在西北大学首先就地质趋势分析方面做了大量工作,后来克伦宾与格雷比尔(1965,第十三章)把这些工作成果加以总结,并提供了一份详细的参考文献目录。

地质趋势分析离不了图,它可以是一维、二维、三维的、或(罕见)甚至是三维以上的。克伦宾(1966 a)指出,二维图是“地质学中的基本通讯工具”。这句话确实不是夸大。图件特别便于表现综合信息,并经常被地质工作者用于此目的。确实,正是这种对图件的熟悉,使趋势分析在地质工作者中广为流行。

地质趋势分析中,统计模型的系数是估计的,方程被表现为等高线图。这种图与地形等高线图的类似,提示我们可用同样的方法绘制这两种图,但必须强调其重要差别。细想一下地质工作者怎样用大平板仪绘制地形等高线图。已故的柯克·布赖恩象考虑基本地质学原理那样仔细地考虑过绘制大平板图的细节,作了象这样的指示:控制点定在坡度改变之处及坡梁上。仅测量点网上有大的空白处的其他控制点。勿试图越出等高线。在野外绘制等高线,因那里可看到地形的细节,必要时还可测量更多的控制点。

遵循这样的指示，地质工作者可按任何的精度要求绘制表示地形的等高线图。各控制点或数据点都是取得有意思的，人的脑子有能力把它们综合成连续的整体。因图件是受过地形与地貌学训练的人在野外所绘制，它比测量人员所制的好，后者是用经纬仪在野外定点，然后在办公室内机械地画等高线。

另一方面，或许除去对于示范性的练习外，趋势面分析中几乎从来不涉及画地形等高线。画等高线的数据反而可能是井或钻孔的数据，或许是来自地表露头的矿物数据或化学分析数据。画等高线的观测值是在这样一些点取得的，从这些点中间一般都看不到天然连续的痕迹。数据点虽可能是按网格分布的，但更常见的是，它们是偶然地分散或集中的。如果竟然有补充观测值，特别是地下观测值，也只能是以相当代价取得的。反之，用大平板仪制地形图，可费力小而恰在需要之处取得补充点。因此适合于在趋势面分析中绘制等高线数据的模型，与适合于绘制地形等高线数据的模型不同。地质工作者虽然承认，即使细节都全知道，也可能需要复杂的曲面来表示它们，但他还是了解，最可行的表示必须包含相对光滑的等高曲线，因数据决不会多到象画地形图所需要的那样多。结果所得的图可以是等厚图，孔隙度图，构造等高线图，矿石品位图，矿物变化图，或者还有其他类型的图。

想一想地质工作者是怎样凭手画非地形数据的等高线的，比如在标准层以上的地下标高的等高线图。正如学生练习画等高线那样，各人画的图不同，每个地质工作者画的等高线图也是不同的。图之所以不同，是由于各地质工作者头脑中有不同的地质模型；最好的图应与最好的模型相一致。例如熟悉洛杉矶盆地的地质工作者，头脑中有断层的模型，通过他的模型画出这地区中的地下数据等高线图，会比所想模型

不适当的人画得更好。

其次考虑可怎样用统计方法绘制这同一数据的等高线。正如地质工作者以地质模型为指导来凭手画等高线那样，地质工作者首先是在头脑中有一个统计模型。区别在于：统计模型必须精确地写成公式，而地质模型则否。这一点之所以要强调，是因为有些研究者曾错误地认为，趋势面分析是“客观的”，而凭手画等高线是“主观的”。相反地，两种方法都有主观和客观的成份。

趋势面分析的一般统计模型是回归分析的一种形式。回归分析中，因变量 w 被表示为一个或多个自变量 x_i 与常数 β_i 的函数加随机起伏 e 。数学上表示为

$$w = f(x_1, x_2, \dots, x_k; \beta_1, \beta_2, \dots, \beta_k) + e,$$

其中 e 服从一种频率分布，例如正态分布。现在“回归”这名称是随意的而非描述性的，不过它在历史上有描述性意义。

在任何回归分析问题中，关于数学函数和 e 的频率分布都有明确的假定。对于单独的一组数据，总有备供选择的假定，必须主观地选择一组为最适当的假定。

一个简单实例

为了使引言这一节的论述更有意义，现在举一个怎样把数据制成等高线图的简单实例。讨论首先联系到地质模型，然后联系到统计模型。例子是西科罗拉多州勒吉里油气田的一组地下观测值(托马斯等, 1945)。数据是在 26 口井中测定的韦伯砂岩(宾夕法尼期或二迭纪)顶部在基准面上的标高(图 9.1)。

概括这些数据的一种方法是按 26 口井中的标高画构造等高线。在地表地质图的观察中，或在由丹佛至盐湖城的商业航线之一上的勒吉里上空的飞行中的观察中，用来指导画

等高线的地质模型是明显的，因构造容易由空中看到。在地面上，由出露的白垩纪曼科斯及梅塞弗德岩层所表现的地质构造为一双倾轴背斜。由于地下岩层和出露岩层基本上是整合的，地下的地质模型也是双倾轴背斜。

对于与地质模型相应的统计模型，光滑¹⁾连续凹函数或凸函数²⁾是适当的。由适合这种要求的许多统计模型中，我们选取形式为

$$w = \alpha + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy + e$$

的二次方程，其中 w 是韦伯砂岩顶部在给定地理点的标高， x 是东坐标， y 是北坐标， α 及 β_1 至 β_5 是参数， e 是假定的随机起伏。当参数被用下述方法进行估计时(九章三节)，方程成为

$$\begin{aligned}\hat{w} = & 2629 - 393.0 x - 968.8 y + 20.7 x^2 \\ & + 153.9 y^2 + 48.2 xy,\end{aligned}$$

其中 $\hat{w}$ 是估计光滑函数值的统计量。为了获得图 9.2 中的构造等高线图，仅需就 和所选 w 值相应的靠得很近的网格点 (x, y) 求此方程的值，为了所选择的等高间距，这些值一律都是百呎数。这分析的复杂性及条件留待九章五节讨论。

趋势面分析的目的

为了不同的目的曾对各种地质数据进行过趋势面分析。这些数据来自火成岩(惠顿, 1962)，沉积岩(克伦宾, 1966；格

1) 就一元函数而论，函数 $f(x)$ 光滑，可简单地了解为其图形上每点处都有不垂直 x 轴的切线——译者

2) 就一元函数而论，设 $f(x)$ 是定义在开区间 I 上的函数， x_0 为 I 上一点。则过点 $[x_0, f(x_0)]$ 的直线 L 有方程 $y = f(x_0) + \lambda(x - x_0)$ 。若对于 I 中每点 x ，曲线上的点 $[x, f(x)]$ 都在直线 L 上的点 $[x, f(x_0) + \lambda(x - x_0)]$ 之上或与之重合，即 $f(x) \geq f(x_0) + \lambda(x - x_0)$ ，则称函数 $f(x)$ 在开区间 I 中凸。如果 $-f(x)$ 凸，则称 $f(x)$ 为凹——译者

图 9.1
 科罗拉多
 州勒吉里
 油田 26
 口井中韦
 伯砂岩标
 高

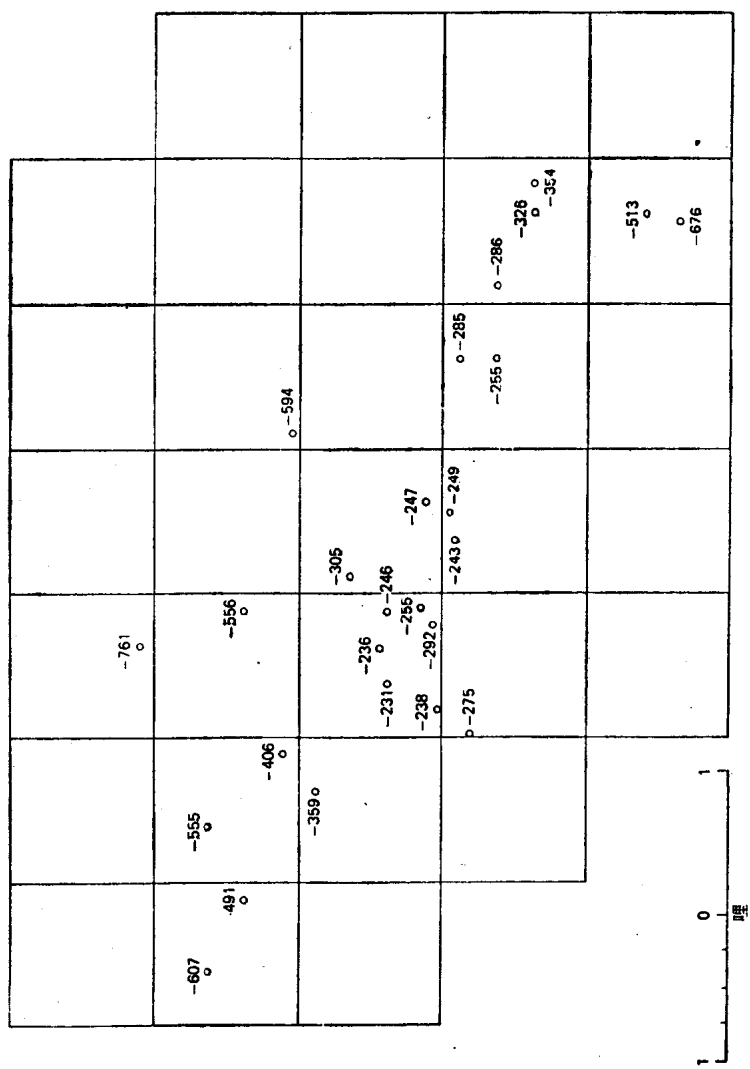
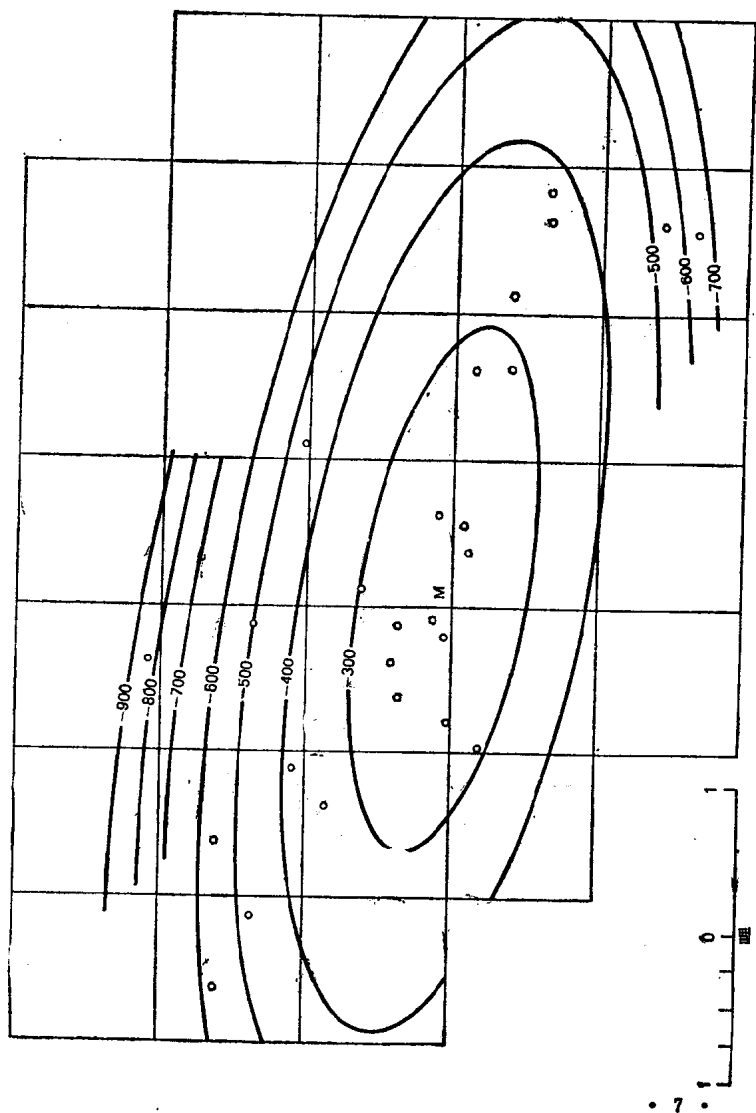


图 9.2
用二次迴
归分析配
置的韦伯
砂岩顶部
构造等高
线图



里菲思,1967),矿山地质(科克与林克,1967),石油地质(哈博,1964)以及其他领域方面的研究。

大多数情况中,趋势面分析被用来概括那些多得一下子不容易看得过来的数据。趋势面分析也容许跟预测的可靠性的定量评价一起,预测在点网内外的值。另一目的是估计趋势面与基准面间或二趋势面间的体积。其他的用途是在估计平均数时使置信范围变窄(这些平均数也许是矿石的平均品位或岩石的平均长石含量)和求得剩余值(它们是数据点处的观测值与预测值的差)。趋势面分析的这些目的及其他目的,以后将在本章中加以探讨。

有许多不是趋势面分析的事。当目的是为所有数据点准确地拟合曲面时,趋势面分析不是用计算机自动为数据绘制等高线的方法。它也不是如有些热衷于新方法的作者所建议的可盲目使用的方法。虽可用各种类型可交换的经纬仪或水准仪绘制地形等高线图,但在计算机技术现有状况下,却不能用各种可交换的计算机程序与计算机绘制趋势面图。更正确地说,趋势面分析的正确目的与计算机程序的性能需要分别评价。

第二节 一维数据的线性回归分析

前节介绍了地质趋势,现在转到正式的分析。本节说明最简单的一种回归,一维数据的线性分析的回归。说明分两部分:供直观了解的描述性分析及适于实际计算的统计分析。

在本章的其余部分,将本节所介绍的原理推广至较复杂的二维及三维的情况。

描述性分析

用一个例来介绍对点拟合直线的描述性分析。设有一条沿矿脉的直的沿脉坑道,某种金属的化验值(w 值)由一端的高值开始,以有规则的直线式向另一端逐渐减低品位。这直线式的,即线性的品位趋势方程是

$$w = \alpha + \beta x + e,$$

其中 x 值是沿坑道的距离, w 是任一点 x 处的化验值, α 是名为截距的回归系数, β 是名为斜率的回归系数, e 是随机误差。

α 与 β 是参数(本书中按惯例用希腊字母代表参数,罗马字母代表统计量)。在举例的沿脉坑道中本来有无限多对的位置点值和化验值,它们是由重复抽取矿样可能取得的;这些观测值组成总体。然而实际上在坑道中只取了几个矿样来确定组成统计样本的值。因此需要新的方程

$$\hat{w} = a + bx$$

——这是与原方程

$$w = \alpha + \beta x + e$$

相应的方程,参数 α 与 β 被换为相应的统计量 a 与 b 。

在沿弗雷斯尼罗矿的矿脉的一条沿脉坑道中逐次抽取矿样,对代表其银含量值的22个点拟合一直线(图9.3及表9.1),可说明上列方程的意义。沿坑道的距离描在水平的 x 轴上;其数与间距为2米的样品位置相应。各样品位置上的银含量值描在垂直的 w 轴上。对这些样品点拟合的直线方程为

$$\hat{w} = 480.1 - 11.74x,$$

其中 $\hat{w}$ 是估计平均银含量, x 是距离。截距 a 为480.1,它是距离为零处的估计银含量;斜率 b 为-11.74,它是距离每增