

地质数据统计分析

上册

〔美〕小乔治·S·科克 著
理查德·F·林克

科学出版社

51.73

地质数据统计分析

上册

〔美〕 小乔治·S·科克 著
理查德·F·林克

王仁铎 刘绶堂 译

赵鹏 大学 校

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。全书分六编，共十七章。上册包括第一至三编，共八章。第一编即第一章绪论。第二编包括第二至六章，介绍了对地质数据统计分析不可缺少的单元统计分析方法，其中包括分布、抽样、推断、方差分析，以及分布和变换。主要目的是讲统计方法，所举地质实例是为了说明方法。第三编一般地讨论地质中的取样和变化性，包括第七、八两章。本册的重点是介绍某些基本统计方法，数学只用到代数与几何，不用微积分。本书可供广大地质工作者参考。对应用统计有关的统计工作者也有一定参考作用。

George S. Koch, Jr. Richard F. Link

STATISTICAL ANALYSIS OF GEOLOGICAL DATA

Vol. 1

John Wiley & Sons, Inc. 1970

地质数据统计分析

上 册

【美】小乔治·S·科克 著
理查德·F·林克

王仁铎 刘绶堂 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：12 1/4

印数：95601—12,320 字数：277,000

统一书号：13041·258

本社书号：1082·13—14

定 价：1.60 元

译 者 的 话

统计方法在地质学中的应用自十九世纪初就逐渐发展起来了。不过，所用的方法都是单元分析的方法，而且计算上受到工具的限制，不能处理大量数据。直到电子计算机出现以后，某些多元统计分析方法才开始成功地应用于地质学，它们当中有：判别分析，趋势面分析和因子分析等。这些方法的引入，使地质学的研究由定性逐渐转入定量。进入七十年代，国外很多地质工作者都在用多元统计分析方法研究他们的课题，涉及范围之广几乎遍及地质学的各个领域，目前已进入到几种方法的综合使用阶段。

在六十年代前，我国只有少数地质工作者使用单元统计方法。六十年代后，特别是在伟大的无产阶级文化大革命的推动下，我国电子计算技术迅速发展，研制了每秒运算百万次的电子数字计算机。在毛主席革命路线的指引下，我国数学工作者和广大地质工作者相结合，深入生产实际，使数理统计的基本知识和方法在广大地质工作人员中迅速普及。在短短几年内，全国各地成立了不少数学地质小组，多元统计分析方法在地质学的各个领域中得到日益广泛的应用。目前应用较多的方法有：趋势面分析，判别分析，回归分析（包括逐步回归分析），因子分析等。

但我们所知，目前还很少出版关于地质数据统计分析，特别是多元分析方面的书籍。为了在广大地质工作人员中普及多元统计分析方法，促进地质科学的定量化研究，同时，也使数学工作者了解多元统计分析在地质方面的应用，我

们将《地质数据统计分析》一书翻译出版，以供参考。

这部书分上、下两册。全书共分六编，上、下册各三编。上册中第一编是绪论；第二编包括第二至六章，介绍了某些基本的统计学概念和单元统计方法；第三编包括第七、八两章，阐述地质学中的抽样和变化性。下册的三编中，第四编阐述多元统计方法；第五编介绍了应用地质学主要是矿山地质数据的统计分析；第六编是讲用电子计算机执行统计分析的。上册的重点是阐述基本的统计学概念和方法，所用的数学范围基本上是初等数学，虽然列出了个别用积分表示的公式，但这对于不熟悉高等数学的读者阅读本书并无妨碍。虽然用了简单的矩阵代数，但对于懂得初等代数的读者来说，学习起来不会困难。下册侧重真实地质数据的分析。在读过了上册的基础上来阅读下册，就可以分出部分精力和时间来考虑具体地质问题的分析了。对于在统计和地质两门科学中只熟悉一门的读者来说，本书的编写方法是很适合的。但在另一方面，本书也有不足之处。例如，有时对频数和相对频数强调不够，使读者在阅读中易于混淆；在举例方面，书中虽给了若干地质实例，但和统计分析在地质学领域中的应用范围相比，则所涉及的范围似较狭窄。尽管如此，在目前普及多元统计分析方法的情况下，本书还是有一定参考价值的。为了读者便于查阅参考文献，书末附有中英人名、地名、矿、矿脉名等对照表。

本书对原书中一些带有资本主义观点和与地质无关的词句、段落作了删改，但也保留了个别举例。翻译过程中我们发现原书有些文字和数字上的错误；对于这些错误，有些是通过查对所引原著予以改正，有些是根据我们的验算或理解作了更正，并加了译者注。1977年，中国文字改革委员会和国家标准计量局“关于部分计量单位名称统一用字的通知”颁

发时，本书已交印制，未及更正。本书虽是为广大地质工作人员参阅的，也可供其他方面的工作者参考。希望读者遵循伟大领袖和导师毛主席关于“洋为中用”的教导，取其精华，去其糟粕，批判地阅读，以促进我国地质事业的现代化。在翻译过程中我们得到中国科学院地质研究所和其他有关单位的一些同志的大力协助，特别是武汉地质学院找矿勘探教研室为我们审阅了初稿，提出了很多宝贵意见，我们在此表示衷心的感谢。由于我们的水平有限，疏漏之处，在所难免，敬希广大读者批评指正。

1977年1月

作 者 序*

所有同数打交道的地质工作者都正式地或非正式地使用统计方法,并从应用统计学的某些知识中得到益处。本书的目的是说明某些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。

我们是为占有数据并想从中得出结论,或其问题的解决需要取得数据和解释数据的人写这部书。我们侧重基本统计方法,并强调这些方法的谨慎应用将产生有效的结果。没有一种统计方法是为了方法本身而提出的,它之所以被提出,只是由于它已被证明对数据分析有用。因为地质学这门科学的内容复杂而性质各异,我们有目的地在统计方法之外加入一些必然涉及的地质论证。可是,其中所用的数学却比较简单,了解它只需用初等代数与初等几何。

本书是为有一些地质训练,但也许不一定有专业经验的读者写的。虽然大多数读者也许是地质工作者,但矿业和石油工程技术人员,地球化学工作者和地球物理工作者,矿物经济工作者及其他的人们也可能对本书感到兴趣。

第一章所概述本书的范围与编排,反映我们在地质学、统计学及电子计算的目前的条件下,对现在最有用的统计方法的选择。本书分为两册。组成上册大部分内容的前几章,把统计学放在第一位,而把地质学放在第二位,因为可把同一种统计方法应用于不同的地质问题。大多数在这一册里所分析

* 作者序中后面的段落作了必要的删节。——译者

的例子,都是为说明特殊统计方法的数值性质而选择的,并是虚构的。下册则是把地质学放在第一位,而把统计学放在第二位,大多数例是真实的,就这些例子把上册所介绍的统计原理加以应用、推广,并在必要时加以改进。书中所引用的数据,没有一处是被全部分析的大批数据,因为这样做就需要一篇专题性介绍,对于一本教科书说来,这就太长了。

虽然举例的数据来自地质学的许多领域,但大多数都来自经济地质学,并有许多是我们亲自分析过的。我们认为,读者从我们处理真实数据的经验中,将会比从对虚构数据或别人的第二手数据的分析中得益更多。但这里存在两种缺陷,首先是对于不太擅长经济地质学的读者,需要作一番努力,才能从所举这些例运用到他们所从事的地质学领域中去。其次,本书对我们自己的工作提到的多一些。这从真实问题的分析中学习,都会弥补这两种缺陷。

本书既是给大学生在正式课程中用的,又是为实际地质工作者写的。我们这本书和 J. C. R. 李的一部应用统计方面的书有密切联系,并采用了他的符号,他的这部书对于自学是很好的。上册后面附有几页符号,我们希望这将使熟悉其他数学符号的读者感到方便。

1970 年 2 月

目 录

译者的话	(v)
作者序	(ix)

第一编 绪 论

第一章 绪论	(1)
第一节 本书的范围和编排	(2)
第二节 问题的解决	(5)
第三节 地质数据的性质	(7)
第四节 作为分析实例用的数据来源	(12)
第五节 模型	(17)

第二编 单元统计方法

第二章 分布	(30)
第一节 观测值	(30)
第二节 经验频数分布	(31)
第三节 直方图	(32)
第四节 理论频数分布	(34)
第五节 概率	(38)
第三章 抽样	(45)
第一节 总体和样本	(45)
第二节 目标总体和抽样总体	(46)
第三节 平均数、方差、标准差和变异系数	(51)
第四节 随机样本	(59)
第五节 抽样分布	(64)
第六节 样本平均数的分布	(71)
第七节 样本方差的分 布 和 χ^2 分布	(77)

第四章 推断	(86)
第一节 总体参数的估计	(86)
第二节 点估计和区间估计	(89)
第三节 司徒顿 t 分布	(93)
第四节 总体平均数的置信区间	(96)
第五节 用置信区间估计矿石品位的例	(106)
第六节 总体方差的置信区间	(111)
第七节 假设检验	(112)
第八节 贝叶斯统计	(129)
第九节 指示	(135)
第五章 方差分析	(139)
第一节 两个以上平均数的比较	(140)
第二节 F 分布	(145)
第三节 单向方差分析	(148)
第四节 单向方差分析的例	(153)
第五节 数据套的方差分析	(158)
第六节 方差分析套的例	(160)
第七节 随机化区组	(163)
第八节 线性组合	(171)
第九节 单自由度	(174)
第十节 用单自由度分析的实例	(186)
第十一节 线性模型	(192)
第十二节 多重比较	(205)
第十三节 基本统计分布间的关系	(212)
第六章 分布和变换	(214)
第一节 几种离散分布	(215)
第二节 对数正态分布	(226)
第三节 变换	(244)
第四节 数据的图形表示法	(249)
第五节 混合频率分布	(257)

第三编 地质中的取样和变化性

第七章 地质取样	(268)
第一节 引言	(268)
第二节 影响取样的地质因素	(269)
第三节 露头的手标本取样	(277)
第四节 钻机取样	(284)
第五节 刻槽取样和拣块取样	(297)
第六节 实际矿物分析	(301)
第七节 其他取样方法	(305)
第八节 地质变化性的范围	(305)
第八章 地质数据中的变化性	(314)
第一节 地质变化性的问题	(314)
第二节 自然变化性	(315)
第三节 取样的偏倚和变化性	(317)
第四节 加工的偏倚和变化性	(318)
第五节 分析的偏倚和变化性	(331)
第六节 实验设计	(340)
附表 1 随机数表	(352)
附表 2 标准化正态分布的百分点	(356)
附表 3 χ^2 分布的百分点	(357)
附表 4 t_m 分布的百分点	(359)
附表 5 F 分布的百分点	(360)
附表 6 司徒顿化全距的百分位	(366)
附表 7 几何平均数的乘数因子	(372)
附表 8 $\rho=0$ 时的 r 的分布的百分位	(373)
部分符号表	(374)
中英人名对照表	(376)
中英地名、矿、矿脉名等对照表	(379)

第一编 绪 论

第一章 绪 论

每个地质工作者都取有数值资料——从岩层收集的手标本数；地层或断层的走向和倾斜；岩石的化学分析，孔隙度测定值；矿石，煤，石油的化验值；等等。而且，每个地质工作者在准备报告，交流其研究结果时，都要概括这些数值资料，把它作为科学处理的一个重要部分。他之所以要概括，是因为通常在地质学科中所取得的数据，都多到不能把逐个数字加以报告；例如地质工作者可能由于图上太拥挤而略去走向和倾角的测定值，或可能把岩石的化学分析或实际矿物分析的数据加以平均。正是他概括的什么以及怎样概括，将在很大程度上决定其报告的价值。

占有大量数据的地质工作者必须使用统计方法，不论是正式地还是非正式地使用。作者的目的是，地质工作者可由有目的地提出的导致明确结论的统计方法使他的思想变得敏锐，并改进其结论的可靠性。这样他就可能避免现今过于经常产生的一种后果——只是把未经解释或解释错误的数值资料表附在他的报告上来作为装饰。

虽然日益大量增多的地质工作者在研究统计学，但学术研究和有益的实际应用间仍有差距。大多数统计学方面的教科书着重这样一些问题，例如从坛子里取出黑的和白的弹子，测定灯泡的寿命，以及数出漏啤酒瓶的数等。很少见到时间和空间关系的问题，而这种问题却是地质上考虑的中心。因

此地质工作者难于将统计学理论转移到他们所直接从事的问题上去。

本书中侧重的是数据分析而不是正规统计学的应用。对于把统计应用在地质上感到兴趣的地质工作者，虽然必须学习一些正规的统计技术，但对他更为重要的是提高鉴别力和判断力。我们与其叙述那些似乎在数学上基础不牢固而计算上不可靠的复杂分析，不如说明如何把精力分配给现实问题。我们的观点大部分是 J. W. 图基(1962，原书第 2 页)的，他写了下面这一段话：

我感到，在其他许多事情中，我的主要兴趣是数据分析。我用数据分析这个词包括：分析数据的程序，解释从这种程序所得结果的方法，设计使分析变得更容易、更精密或更精确的数据收集方法，以及所有用来分析数据的计算机和（数理）统计的结论。大部分数据分析是推断性的，……但这些推断仅是部分的数据分析而不是它的全体。大部分数据分析是清楚的、明显的指示，不过我们由原始数据的简单而直接的考查却不能理解这些指示。但这些也仅是部分的数据分析，而不是它的全部。数据分析这一术语在这里被引伸得超出其语言学范围之外了¹⁾，它的某些部分是在下列意义上的分配，即某些部分的数据分析指导我们如何在精力和在观测、实验或分析中其他有价值的考虑上进行分配。数据分析的范围比推断，清楚的程序或分配的范围更大而且更加多变。

第一节 本书的范围和编排

本书是为想把统计应用于地质的人员编写的，所以假定我们的读者懂得一些地质学，但不必懂得什么统计学。本书从被充分扩大到能应用于地质资料工作的基本统计原理讲起。

1) “分析”(analysis)这个词在英文中有两种含义：分析的行动和分析结果的说明。——译者

没有企图纵览应用统计学，或对于把统计应用在地质方面的文献加以概括。

虽然统计方法的讨论从基本原理开始，有时却以不包括在基本统计学书中的专门化程序结尾。本书只介绍为解释地质数据和说明地质问题所必需的统计方法，并计算了许多大都以真实地质数据为根据的数值实例。给出了简单的代数论证，但需要微积分的复杂证明则被略去。即使是懂得统计学的读者，为了熟悉我们的观点和符号，浏览一下作为基础的前几章也是有益的。

上述一切都属于本书的范围。下面几段讲到一些本书未加论述的有关论题，并提示说明这些论题的著作。

因为我们讲的是精选的统计问题，即使在初等的水平上，也不打算对统计学作概括的说明。在许多好的统计学教科书中，最适合地质工作者的有两部，它们是 J. C. R. 李的书（共两卷，1964 年版的卷 I 和 1964 年版的卷 II）以及狄克逊和梅西（1969）的书。李的书（特别是卷 I）很适合自学。它是为专业人员写的，并为数学知识有限或生疏的读者非常详细地讲解了统计学。只要读过这两部书中的一部或全部并做了习题，任何人都将获得有助于地质工作者的良好的基本统计学基础。

我们不讨论数理统计，它是许多书的主题，例如赫尔（1962 和 1966）的两本初等统计的书及威尔克斯（1962）的一本高等的书。对于理论统计学的中心论题——概率，我们这里只简短地加以讨论；对它感兴趣的读者可参阅莫斯特勒等（1961）的一本初等教科书及费勒的一本高等的书（卷 I，1957¹⁾和卷 II，1966）。

1) 前九章有科学出版社 1964 年出版的中译本。——译者

关于地质方面的统计学有五本很好的书，它们涉及另外一些本书未提到的课题。米勒和卡恩(1962)对直到1962年的文献作了很好的提要，从此以后，发表的论文之多，已使一本内容完备的评论杂志几乎不能再容纳其他什么了。除评述文献外，米勒和卡恩还讨论诸如概率，概率密度函数和计量古生物学等，我们对所有这一切都几乎未加注意。克伦宾和格雷比尔(1965)侧重统计模型的公式化，其大多数实例取自沉积学和石油地质学。黑曾(1967)在美国矿业局会刊第621号报导了采矿工艺，特种取样法以及矿石和岩石粒度分布。格里菲思(1967)讨论了研究沉积物的统计方法，但他的书有更广泛的应用，包括本书中未提到的统计方法，并提出关于地质科学远景的意见。最后，哈博和梅里亚姆(1968)侧重地层分析的由电子计算机来执行的统计方法。

本书分两册，共六编，每册三编。第一编即这章绪论。第二编包括第二至第六章，解释并通过实例说明单变量分析的单元统计方法；仅引入足够的地质数据来帮助解释统计学。第三编是第七和第八章，提出地质取样和地质变化性。第三编中强调数据解释这个难于对付的问题，因为它属于地质学，因此既注意数据分析的统计学问题，又注意它的地质学问题。第四编由下册中第九至十一章组成，叙述两个或两个以上变量的分析的多元统计方法。第五编包括第十二至十六章，讲应用地质学主要是矿山地质学中数据的统计分析，这些章讲的是自然资源的勘探及其评价，通过运筹学作出决策，以及某些专门化抽样方法和数据分析。最后是第六编，即第十七章，讲应用电子计算机来执行统计方法。

第二节 问题的解决

就地质研究来说,统计学很可能对它有益的是集中解决一个具体问题。简单地概括起来说,好的指导如下:有目标,预订计划,并仅采用将导致该目标的任何有效的地质和统计的方法。

首要的是研究人员应仔细思考所要从事研究的问题:范围、限制、细节和目标。然后他才可以适当方法建立被称为模型(一章五节)的那类假设,有目的地收集数据来检验这项假设,并在所述可靠性范围内解决这个问题——也许在过程中还要借助于初步结果来改进其模型,然后在必要时再收集补充数据来改进和验证其结论。这样,他就既不会被大量数据所淹没,又不致收集不到有关数据。他就能在成本(不论是以金钱计还是以时间计,或二者并用)的范围内得到一个具体问题的有效解决。

为一个目标奋斗成功的例子,是第二次世界大战后美国政府和工业地质工作者在美国发现大的新铀矿床所提出的勘探计划。许多新铀矿是在理论地质学认为不利的环境中通过系统找矿发现的。有意思的是注意到,当主要是为找矿而收集的数据后来被用于次要目的,例如地层或地化研究时,发现这些数据并非完全合适的。如果回忆一下,有些研究人员本来希望他们已经针对不同的新问题的研究,收集了新的数据,可是仍是旧的数据。

在问题的解决中,与其只束缚在传统地质学,不如综合运用可想象到的任何和所有科学及工程领域中有助于问题的解决的方法。正是在这种精神的支配下,本书包括有若干学科中的方法。克劳德(1964)以类似的精神写道:

现今地球科学在哲学上最独特并最令我本人满意的，是古典科学中有用的部分与先进科学中最令人振奋的方面的融合……。当新设备，新测量法和改进的计算机装置为分类和快速定量评价方法提供了各种基础，并在有些情况下提供了更重要的基础时，全部系统科学，这里我是在广义的分类和解释中使用系统这个词的，都处于动荡状态中——这不仅对矿物学真实，而且对古生物学和岩石学也真实……。

如果要列举最一般地刻画现代地球科学特征的东西，我们也许就会包括：

1. 传统研究方法逐渐停滞不前。
2. 只要可能做到，定量地表示观测结果和结论的趋势日益增大。
3. 与其他科学的相互作用正在增长。
4. 对作为一种交流形式的数学，以及对物理学、化学和生物学……的熟悉程度的假定和要求日益提高。
5. 高度复杂精密的仪器能增强我们在所有学科领域中的分辨能力，使我们不仅在新的领域中，而且在一度接触过的领域中进行新的观察和发现。

堪萨斯地质调查所副所长威廉·汉布尔顿(1966)特别对国家地质调查工作持有一种类似的观点。他写了如下的话：

至于现代调查计划，可以说它是由国家和它所影响的地区的社会、经济和政治问题中的变化性，紧迫性和复杂性来表示其特征的……。百年前，我们机构的目的是调查国家的矿产资源。今天，它是经济发展的一种手段；它的研究服务于促进矿产工业的活动；它的计划，目标和方案都以革新为其特征。若干年来我们为解决基本地质问题从事于包括计算机技术在内的计划……。过去几年我们使地球物理学、统计学、石油工程学、矿山工程学、地质学和计量经济学中的一些人陷入在发展区域经济分析方法的研究中。

换言之，我们正越来越多地在和矿产工业经济中的系统分析和运筹学的整个领域打交道。我们正投身于城市发展和环境地质研究，它们涉及到环境卫生，交通运输，土地利用以及城市和地区的规划……。

大多数我所提到的研究都强调改变和新的方向。所用的方法包括