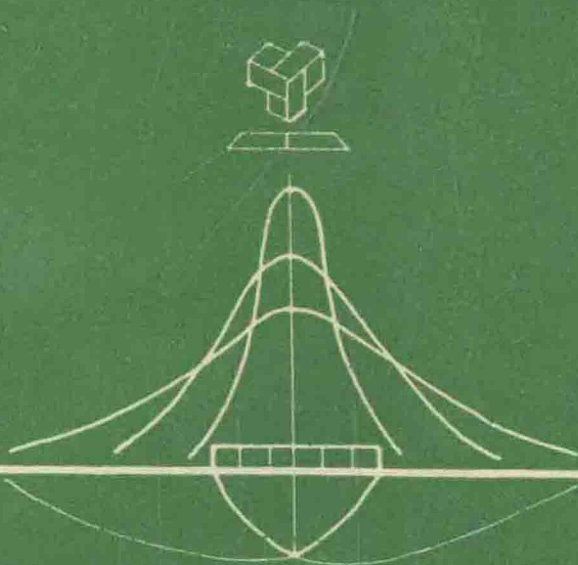


岩土工程勘察数据 统计分析

马良驹 袁灿勤 编著



南京大学出版社

岩土工程勘察数据统计分析

马良驹 袁灿勤 编著

南京大学出版社

1991·南京

内 容 提 要

本书从岩土工程勘察工作的实际需要出发,主要介绍了岩土工程勘察常用的数据整理、计算、分析、判断的数理统计方法。全书共六章,包括概率统计知识、岩土参数的单元统计计算、方差分析的应用,以及多元统计分析的回归分析、正交试验设计、判别分析等。本书除概率统计的基本知识外,没有过多专门介绍概率统计的系统理论,而是把它们有机地结合到具体的统计分析的各种方法中。

本书语言通顺,阐述详尽易懂,并附有大量结合岩土工程勘察方面的举例和例题,便于从事岩土工程勘察、工程地质等专业工作的技术人员学习参考,也可作为上述专业的大专院校学生的参考书。

岩土工程勘察数据统计分析

马良驹 袁灿勤 编著

*

南京大学出版社出版

(南京大学校内)

江苏省新华书店发行 镇江前进印刷厂印刷

*

开本787×1092 1/32 印张10.125 字数227千

1991年1月第1版 1991年1月第1次印刷

印数1—2500

ISBN 7-305-00994-6

定价 4.80元

前 言

近年来,岩土工程勘察的理论和方法得到很大发展,新的测试手段和高精度的测试仪器以及计算机技术在工程实践中不断被采用,大量的测试数据给我们提供了丰富的有价值的岩土物理力学性质方面的各类信息,这就给岩土工程勘察从过去的以定性为主向以定量为主的研究转变提供了有利条件,这种转变的客观要求,在历次规范的修订和重订中可以得到充分的体现。

作为岩土工程勘察对象的岩、土、水是地质历史过程的产物,它们的成因、成分、性质是极其复杂且互相联系、制约和互相影响的,具有很大的不确定性,在很多方面,我们还无法从机理上用常规的数学方法揭露其本质规律,而只能依靠概率统计方法来研究它们的统计规律。因此,到目前为止,概率统计分析是研究大量岩土工程勘察数据可行而有效的方法。

有关概率统计的著作可谓不少,但其内容或者侧重于概率统计的经典理论和系统阐述,或者与本专业联系甚少,不能紧密结合岩土工程勘察实践。

我们觉得,当前应该有结合岩土工程勘察实际的数据统

计分析的这类书,以供广大从事这方面工作的人员参考。

有鉴于此,作者力图使本书做到:

1. 尽量避开概率统计中某些纯理论的严格推导论证,对必不可少的内容也尽量做到深入浅出,通俗易懂;

2. 在介绍岩土工程勘察常用的统计分析的具体方法的同时,重视使读者树立正确的统计思想,对于使用某一方法的诸多细节作较详细的介绍,以便于读者实际应用。

本书共六章,第一、二章由袁灿勤执笔,第三、四、五、六章由马良驹执笔。

最后,特别需要指出的是,由于作者水平所限,成书时间仓促,不妥或错误之处肯定很多,恳请读者不吝指正。

作 者

一九九〇年九月于南京建筑工程学院

目 录

第一章 概率统计基本知识	1
第一节 引言	1
第二节 随机现象和随机事件	3
一 随机现象	3
二 随机事件	4
第三节 概率与频率	6
一 概率的古典定义	7
二 频率与概率的统计定义	7
第四节 随机变量及其概率分布	10
一 随机变量	10
二 离散型随机变量及其概率分布	11
三 连续型随机变量及其概率分布	13
四 正态分布简介	16
第五节 抽样调查的概念	28
一 总体、个体与样本	29
二 随机抽样	31
三 抽样误差	34
第二章 岩土工程勘察数据单元统计计算	37
第一节 数据的初步整理	37
一 不分组的列表方法	38
二 频数、频率分布表	40
三 数据散点图	43
四 频数、频率分布图	44

第二节 统计特征值	50
一 位置特征值	50
二 离散特征值	61
三 分布特征值	75
第三节 岩土工程勘察数据的波动与离群值	81
一 误差对数据波动的影响	82
二 岩土自然变化性对数据波动的影响	83
三 离群值的概念及舍取	84
第四节 岩土特性指标的统计计算	90
一 问题的提出	90
二 确定岩土参数可靠性估值的理论依据	91
三 岩土参数可靠性估值的确定	99
第五节 岩土抗剪强度参数的确定	104
一 直剪试验	105
二 三轴剪切试验	109
第三章 回归分析	116
第一节 回归分析概述	116
一 变量间的两种关系	116
二 回归分析及其在岩土工程勘察中的应用	118
三 回归分析类型	120
第二节 一元线性回归分析	122
一 一元线性回归方程的建立	122
二 一元线性相关系数	128
三 一元线性相关系数的显著性检验	133
四 预测与精度分析	136
五 一元线性回归方程的近似求法	147
第三节 一元非线性回归分析	150
一 回归曲线的选配	151
二 化曲线回归为直线回归	158

三 其它问题	160
第四节 多元线性回归分析	170
一 多元线性回归方程的建立	171
二 多元线性相关程度的检验	182
三 复相关系数的显著性检验	184
四 多元线性回归方程的精度分析	185
五 多元线性回归分析的其它问题	186
第五节 最优回归方程及逐步回归分析	197
一 问题的提出	197
二 回归方程最优化方法	198
三 逐步回归分析简介	202
第四章 方差分析	203
第一节 方差分析概述	203
一 什么是方差分析	203
二 方差分析的基本原理	206
第二节 单因素方差分析	210
一 单因素方差分析的意义及有关计算	210
二 单因素方差分析的步骤与方法	213
第三节 双因素方差分析	216
一 双因素方差分析的意义	216
二 非重复试验的双因素方差分析	217
三 重复试验的双因素方差分析	227
第五章 正交试验设计	233
第一节 正交试验设计概述	238
一 试验为什么要设计	238
二 正交试验及正交表	240
第二节 不考虑交互作用的正交试验设计及成果分析	244
一 设计表头,安排试验方案	244
二 试验结果分析	246

三 混合型正交试验	255
第三节 考虑交互作用的正交试验设计	257
一 交互作用的概念	257
二 试验方案的安排	258
三 试验结果分析	262
第六章 判别分析简介	265
第一节 地质分类及判别分析	265
一 传统的地质分类	265
二 判别分析的基本概念	267
第二节 两组线性判别分析	268
一 判别分析的费舍准则	268
二 向量的线性组合	270
三 判别函数的计算	272
四 两组线性判别方法	276
五 判别函数的显著性检验	280
六 两组线性判别分析举例	281
七 判别分析讨论	288
主要参考文献	291
附表	293

第一章 概率统计基本知识

第一节 引 言

岩土工程勘察数据统计分析,是运用概率论和数理统计的有关理论和方法,对岩土工程勘察工作所获取的大量数据,进行科学的整理、检验、分析、判断,以达到正确评价岩土工程勘察成果和为工程设计、施工提供准确可靠的设计参数的目的。

概率统计是概率论和数理统计的合称,它们是两个紧密联系而又有所区别的数学分支学科,前者侧重于对随机变量一般规律的理论研究,后者侧重于对现实世界中随机变量的收集、整理、检验、分析与判断的具体方法。因此可以认为前者是后者的理论基础,而后者又是前者的实际应用。

由于现实世界中随机变量的普遍性,从而决定了概率统计应用的广泛性,它几乎应用于认识自然和改造自然的一切领域,在国民经济各个部门,如工业生产、农林渔业、气象水文、医疗卫生、交通运输、公共事业等都需要使用概率统计方法解决有关问题。

在岩土工程领域内,由于新的勘察试验手段的不断涌现,现代化高精度的试验、测试仪器的广泛使用,以及计算技术迅猛发展,促使地质学从过去的以定性研究为主向量化研究

发展。这种转变必然向我们提出一个很现实的问题，即勘察所获得的大量数据如何加工处理，从中发掘出规律性的东西？结论是：数学是一个有力武器，它在解决地质问题的量化方面是大有作为的。因此数学与地质学的紧密结合，从而产生了一门新的边缘学科——数学地质。

数学地质是一门新兴的学科，它的发展也只是一二十年的历史，它正在不断的发展和完善之中。概括地说，数学地质是以运用数学各分支领域的成果，解决地质学中的理论和实际问题为己任的，由于地质体和地质现象大多都具有一定程度的不确定性，因此以研究随机变量为对象的概率统计分析，尤其是多元统计分析，在数学地质中占有重要地位。

综上所述，所谓岩土工程勘察数据统计分析，就是运用概率统计的基本原理和方法，解决岩土工程勘察中所获取的大量数据的处理问题，是数学地质学的重要组成部分。学习一点这方面的知识，对于从事或即将从事岩土工程专业的人员来说是很必要的。概率统计是一门理论性很强的数学分支学科，这方面有大量的参考书籍可供参考，在本书中不准备涉及过多的这方面的理论，而着重于在岩土工程勘察方面的应用的具体方法，特别是如何正确运用概率统计思想，解决岩土工程勘察中的有关问题，是学习本课程的重要目的之一。

在介绍岩土工程勘察数据统计分析的有关内容之前，有必要先简单介绍概率统计中有关的几个基本概念，对于未学过概率统计的人，可作为学习以下章节的预备知识，对于学习过概率统计的人，也可作为一个简单的复习。

第二节 随机现象和随机事件

一、随机现象

自然界一切现象,按照统计学的观点可分为两类,一类是确定性现象,一类是非确定性现象。

所谓确定性现象,是指在一定条件下,必然会发生或必然不发生的现象。例如水在标准大气压下,温度达到 100°C 必然沸腾;同一种电阻片,在相同的外界环境中,电阻片的应变增量与其电阻值的增量成正比;砂土试样在增加垂直压力的条件下,抗剪强度必然增大;压力水头高于地表的承压含水层,当揭露其隔水顶板后,地下水必然喷出地表等等,这些现象都是确定性现象。

所谓非确定性现象,是指在一定条件下可能发生也可能不发生而不能预先确定的现象,又称随机现象。例如某条河流今年的洪水水位是否会超过当地的警戒线;某地区潜水含水层在某一时刻的水位标高;测试某土层的标准贯入击数等等,这些现象在它们各自的条件下能否发生以及发生何种结果,都是预先不能确定的。

确定性现象之所以能预先确定,是因为这类现象的发生或发展是受内部某种确定的规律所完全支配,我们掌握了这一规律从而能预定它的发生以及发生何种结果。确定性现象的这种内部规律称为“动态性规律”。随机现象则不然,它不受或不完全受动态规律所支配,而是在一定程度上受无法控制的偶然因素的影响,因而表现出随机性。那么随机现象是否就没有任何规律可循呢?完全不是。恩格斯说:“被断定为必

然的东西,是由纯粹的偶然性构成的,而所谓偶然性的东西是一种有必然性隐藏在里面的形式。”随机现象正是一种内部隐藏着必然性,而在外部以偶然性的形式表现出来的现象。换言之,随机现象的偶然性仅仅表现在这一现象的个别的少量的重复中,而必然性则隐藏在这一现象的大量重复中。当我们重复试验或考察一个随机现象时(这种试验或考察必须在相同的条件下重复),那种个别的表面的偶然性,就被内部隐藏的必然性所代替了。随机现象的这种规律称为“统计性规律”。很明显,要发掘这种规律,其先决条件是能够对随机现象进行大量的重复试验或考察(在数理统计中称为随机试验)。例如对某一岩土层的某种物理、力学指标的多次测试;对某一地区的岩、土、水样品的某种成分多次取样分析等等都可视为随机试验。自然界随机现象之所以能被人们所认识掌握,正是因为它们存在这种统计规律的缘故。概率统计就是研究随机现象统计规律的科学。

需要指出,确定性现象与随机现象并不是一成不变的,两者之间在一定条件下是可以转化的。例如一个确定性现象当其受某些偶然因素的影响,或由于测量、试验而产生随机误差时(一般这是难免的),也会给这一现象带来一定的随机性。反之,一个随机现象如果能够准确地掌握了它的动态规律并能够严格控制其外部的偶然性影响因素(一般是很难完全做到的),也会使这一现象在一定程度上转化为确定性。了解这一点对我们自觉地有效地进行勘察数据的统计分析是有益的。

二、随机事件

概率统计不是笼统地研究随机现象,而是着眼于具有实际内容的随机事件。任何现象必定以某种定性或定量的结果

表现出来,在一定条件下,现象所表现出来的具有具体含义的结果称为事件。在一定条件下某一结果可能发生也可能不发生而预先无法确知的事件,称为随机事件(通常简称为事件)。例如前述的随机现象中,测试某一土层的某种物理、力学指标时出现误差的现象,这是一个随机现象,如果只限于研究误差的正负,那么出现正误差是一个随机事件,出现负误差又是一个随机事件;某土层的标准贯入击数,就其无法预先确定这一点来说,这是一个随机现象,而击数每一个可能取得的结果,则是这一随机现象中的一组随机事件;潜水位例子也是如此。

从上述例子中不难看出,一个随机现象总是由若干个事件所构成,一个随机现象包含多少事件,取决于随机现象的性质和我们主观的研究范围与目的,例如在前述误差的例子中,如果只限于研究误差出现的正负,则这一随机现象只包含两个事件,其事件数是有限的、可数的;如果研究误差出现的各个可能值,则这一随机现象所包含的事件将是充满某一区间的所有实数值,事件数是无限的、不可数的。在前述的标贯击数的例子中,如果只限于研究击数是否会超过某一定值,则这一随机现象也只包含两个事件,其事件数也是有限可数的;如果研究击数出现的各个可能值,则这一随机现象所包含的事件数将是全部正整数系列,事件数是无限的、可数的。

此外,根据事件的相互关系,在概率统计中常把事件区分为以下几种情况:

互不相容事件 设事件组 $A_1, A_2, \dots, A_n$ (n 为有限可数的), 其中任意两个事件不可能同时发生, 称它们为互不相容事件, 也称互斥事件。例如射击比赛中, 击中 $0, 1, 2, \dots, 10$ 这11环的事件就是互斥事件。

等可能事件 由于对称等原因,事件发生的可能性都相等的事件,称为等可能事件或等概率事件。例如投掷一枚骰子出现 $1, 2, \dots, 6$ 点的事件就是等可能事件。

完备事件组 在随机试验中,它的每一个可能出现的结果都是一个事件,它们是这个试验的最简单的事件,这些简单的事件称为基本事件,如果一个基本事件组 $A_1, A_2, \dots, A_n$ 即是互斥事件而且每次试验必有也只有一个基本事件出现,这样的事件组称为完备事件组。上述投掷骰子的例子就是一个完备事件组。

不过岩土工程勘察数据统计分析所涉及的事件,基本都属于非互斥事件和非等可能事件,因此也大多不属于完备事件组,我们在此提及上述有关概念,只是为了在概率统计方面建立一个清晰的概念而已。

第三节 概率与频率

我们考察一个随机现象,不仅需要知道在一定条件下某一事件可能发生或不可能发生,重要的是需要知道它发生或不发生的可能性大小,只有这样才算是掌握了随机现象的内在规律,并利用这一规律解决生产实际问题。例如修建大型水坝时,需要根据水库上游的特大洪水位设计坝体高度,特大洪水位的出现是一个随机现象,如果我们仅仅知道它可能发生或可能不发生,是没有意义的,我们需要知道它发生的可能性有多大,是十年一遇、百年一遇,还是千年一遇,这就是事件的概率问题。因此通俗地讲,概率就是衡量事件发生的可能性大小的一种尺度。下面介绍概率的两个基本定义。

一、概率的古典定义

有些随机试验所包含的事件具有某种特定的概率特征，据此可以把它们归纳成若干特定的概率型式，简称为概型，其中一种较简单也比较常见的概型称为“古典概型”。如果随机试验所包含的事件是一个等概率的完备事件组，就表明这个事件组是符合古典概型的。事件组如果属于古典概型，概率的定义可描述如下：

设 $A_1, A_2, \dots, A_n$ 为一个等概率的完备事件组，则事件 A 的概率 $P(A)$ 可用下式表达：

$$P(A) = \frac{m}{n} \quad (1-1)$$

式中 m —— 属于事件 A 的基本事件数；

n —— 完备事件组的基本事件数。

上式就是概率的古典定义的数学表达式，只要事件符合古典概型的条件，并确定了 m 和 n ，由式 (1-1) 即可计算任意事件的概率，而不需要对该随机现象进行重复试验或考察。不过确定 m 和 n 有时不像某些简单的情形那样，可以直观的加以确定，在复杂的情形下，要根据排列组合的原理计算。但正如前面所提到的，岩土工程勘察数据统计分析的对象绝大部分不符合等概率的完备事件组，因此也就不符合古典概型。下面着重介绍概率的另一定义。

二、频率与概率的统计定义

在介绍概率的统计定义之前，先介绍频率的概念，因两者有密切的联系。设事件 A 在同一条件下进行了 N 次重复试验或考察，事件 A 出现了 f_A 次，则事件 A 的频率 $W(A)$ 可用下式

表达:

$$W(A) = \frac{f_A}{N} \quad (1-2)$$

式中 N ——随机试验的次数;

f_A ——在 N 次试验中事件 A 出现的次数。

为说明频率的性质,让我们先举一个在统计中经常引用的掷硬币的例子。设任意投掷一枚硬币,出现正面为事件 A ,根据式(1-2)可知,出现正面的频率 $W(A)$,既与出现正面的频数 f_A 有关,又与抛掷的次数 N 有关。现设

抛掷一次($N=1$),则

$f_A = \{1, 0\}$,相应的 $W(A) = \{1, 0\}$,即事件 A 的频率非1则0,频率的波动最大。

抛掷二次($N=2$),则

$f_A = \{2, 1, 0\}$,相应的 $W(A) = \{1, 0.5, 0\}$,即事件 A 的频率在1, 0.5, 0三个值上波动,而且从经验中可知,三个频率值中取得0.5的可能性大于其它两个值。

抛掷五次($N=5$),则

$f_A = \{5, 4, 3, \dots, 0\}$,相应的 $W(A) = \{1, 0.8, 0.6, \dots, 0\}$,即事件 A 的频率在1, 0.8, 0.6, ..., 0六个值上波动,而且越趋于中间的值的可能性越大,越趋于两端的值的可能性越小。

这说明抛掷次数越少,事件 A 的频率波动越大,反之抛掷次数越多,频率的波动越小,这就是频率的不稳定性。设若抛掷的次数更多,事件 A 的频率将如何?历史上曾经有人作过如下的抛掷实验(表1-1)。

实验表明,随着抛掷次数的增加,频率将逐渐趋于稳定,因此有理由认为,当抛掷次数趋于无穷时,频率将稳定在某一