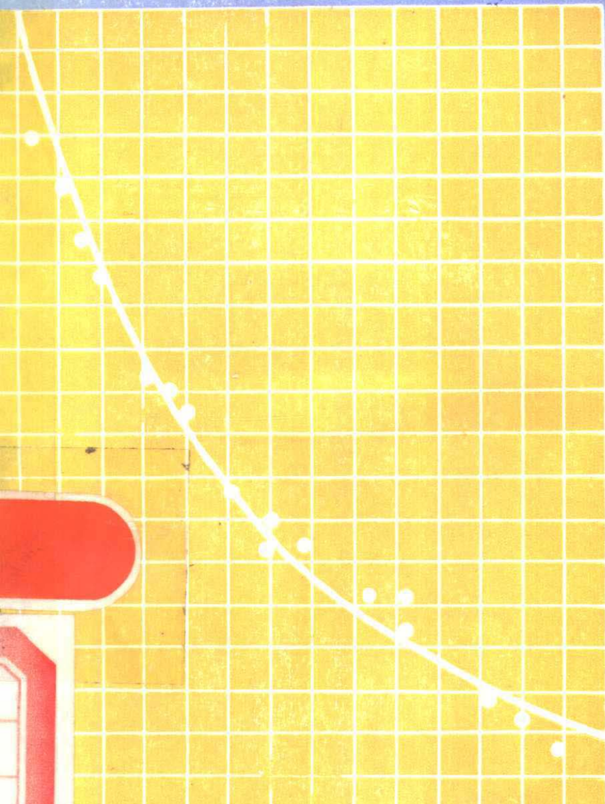
沙庆林 编著

# 观测试验资料的数学加工法

(第三版)



人民交通出版社

Guance Shiyan Ziliao de

# 观测试验资料的

Shuxue Jiagongfa

# 数学加工法

（第三版）

沙 庆 林 编著

人 民 交 通 出 版 社

## 内 容 提 要

本书围绕试验资料的统计处理，着重从应用方面介绍误差理论，试验结果的精度和表示方法，几种概率分布的具体应用，最少试验数量的确定方法，经验公式的选择及回归分析方法等。本书具有如下特点：内容较新，份量比第一版增加了两倍；切合实用，实例多达160个，便于读者对内容的理解；由于书中还简单介绍了如何利用电子计算器及PC-1500袖珍电子计算机进行运算的方法和回归分析电算程序，便于读者提高工作效率。

观测试验资料的
数学加工法
（第三版）
沙庆林　编著
责任编辑：武崇理
封面设计：梁毓英
插图设计：裘　琳
技术设计：张义华
责任校对：梁秀清
人民交通出版社出版
新华书店北京发行所发行
各地新华书店经售
人民交通出版社印刷厂印
开本：787×1092 $\frac{1}{32}$　印张：17.5　字数：387千
1960年8月　第1版
1988年4月　第3版　第3次印刷
印数：17,301—20,350册　定价：4.75元

# 前　言

笔者于1960年曾编写了《观察试验资料的数学加工法》一书，1975年再版时，没有来得及进行修改。近十多年来，国内出版了多种版本的数理统计书籍，使得数理统计方法逐渐成为各个领域的科研人员和观测试验人员处理试验数据的有力工具。数理统计方法在道路工程中的应用也日益增多。生产实践和科学试验的深入发展，要求更广泛和深入地应用数理统计方法，以便经济合理地进行观测试验，科学地处理试验数据。为适应当前读者的需要，这次修订，内容作了较多的补充，篇幅比第一版增加了约两倍，对原内容也进行了修改。

对试验数据的科学处理，通常称做统计处理。因此，本书本应改名为《观测试验资料的统计处理》，但考虑到上述书籍自1960年第一版以来，重印几次，读者较为熟悉，故仍用原书名（只将“察”改为“测”）。

本书仅简略地介绍概率论中的几个基本概念，也没有全面阐述数理统计的各个方面。对于数理统计中的一些理论和公式的推导，一般未作详细的说明，只是着重从应用方面来阐述一些原理和公式，并用大量实例来说明这些原理和公式的应用。对理论推导感兴趣的读者，请参阅概率论和数理统计专著。

为便于叙述并便于读者理解，全书共分十八章。在第一章中，根据国际标准介绍了书中所用的术语和符号。在第五

章中，除着重讨论试验精度和试验结果的表示法外，还着重讨论了一般函数误差的表示法。鉴于正态性检验的重要性，书中单列一章对其进行了较详细的介绍。在多种情况下，如何根据试验精度和要求确定试验的最少试验数量，对实际工作和数据的可靠性具有十分重要的意义。因此，在第十三章中较详细地论述了最少试验数量的确定问题。数据统计处理中的全部计算，利用现代化的计算工具，都可以既快速又准确地得出结果。为了帮助读者利用这些工具，在第十八章中，以科学电子计算器SHARP EL-5100为例，举例说明如何利用科学电子计算器进行快速运算；同时，以PC-1500袖珍电子计算机为工具，用BASIC语言编制了一元线性回归方程、曲线方程以及二元线性回归方程的计算机程序，供读者参考应用。通过变量变换，可以使很多曲线关系式变换成直线关系式，然后进行回归分析。鉴于此问题的重要性，书中专门讨论了经过变量变换后所得关系式的剩余标准差和预报值的波动界限问题。

书中所用到的全部概率分布表和其他数表都汇集在附录二中。本书编写过程中的参考文献列在正文的后面，文中不再一一注明。

对本书中可能出现的缺点和错误，欢迎读者批评指正。

# 目 录

# 第一章　术语和符号

本章仅涉及本书中所用到的术语和符号，以方便读者阅读书中的具体内容。本章所用的术语及其定义和符号是完全参照国际标准 ISO 3534—1977（E/F）。

## 第一节　术语及其定义

1.随机变量或变量：一个可以取一特定数集中任何值的，而且与其相应有一概率分布的变量（或称变数）。

只有孤立值的随机变量称为“离散的”。可以取有限区间或无穷区间的所有值的随机变量称为“连续的”。

2.随机变量的概率分布：一个确定随机变量取任一给定值或属于一给定数集的概率的函数。

随机变量变化的整个区域的概率等于1。

3.分布函数：是函数的一种，即对于 $x$ 的每一个值，此函数给出随机变量 $X$ 小于或等于 $x$ 的概率。

$$F(x)=P_r[X\leqslant x]$$

4.连续随机变量的概率密度函数：即分布函数的导数（当它存在时）。

$$f(x)=\frac{dF(x)}{dx}$$

（$f(x)dx$ 是“概率元素”；$f(x)dx=P_r[x<X<x+dx]$）

5.离散型随机变量的概率：设 $x_i$ 是离散随机变量 $X$ 所取的一个值，则概率 $p_i$ 是：

$$p_i = P_r\{X = x_i\}$$

6.相关：两个或几个变量间的相互依赖关系，在这一关系之中包含有随机部分。

7.概率分布的分位数：设 $p$ 是 0 和 1 之间的一个数，$p$ 分位数是随机变量的一个值，它使得分布函数在该点的值等于 $p$，或是从小于或等于 $p$ 的值“跳”到大于 $p$ 的值。

分布函数在变量的两个相邻可能值间的整个区间有可能都等于 $p$。在这种情况下，此区间中的任一个值都可以作为 $p$ 分位数。

8.中位数：概率分布的 0.5 分位数。

9.随机变量的数学期望（平均值）：

1)对一个取值 $x_i$、概率为 $p_i$ 的离散型随机变量 $X$，其数学期望定义为：

$$E(X) = \sum p_i x_i$$

对 $X$ 可能取的所有值 $x_i$ 求和。

2)对一个具有密度 $f(x)$ 的连续随机变量，其数学期望定义为：

$$E(X) = \int x f(x)\,dx$$

积分是对 $X$ 的变化区间的所有值取的。

随机变量的数学期望与其概率分布的数学期望没有什么差别。

10.标准化变量：数学期望等于0，标准（偏）差等于1的随机变量。

注1：如随机变量$X$的数学期望等于$E(X)$、标准差等于$\sigma$，则与其相应的标准化变量是

$$\frac{X-E(X)}{\sigma}$$

标准化变量的分布称为“标准分布”。

注 2：标准化变量的概念可以推广为“化约变量”。化约变量是用别的位置参数和尺度参数来定义的。

11.正态分布（拉普拉斯-高斯分布）：连续随机变量 $X$ 的概率分布，若 $x$ 是任一实数，其概率密度是

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$-\infty<x+\infty$$

式中：$\mu$——正态分布的数学期望；

$\sigma$——正态分布的标准差。

12.标准正态分布：标准正态变量的概率分布。

对一个参数为 $\mu$ 和 $\sigma$ 的正态变量 $X$，其标准正态变量是

$$Z=\frac{X-\mu}{\sigma}$$

分布变量 $Z$ 的概率密度是

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

$$-\infty<z<+\infty$$

13.$\chi^2$（卡方）分布：独立的标准正态变量的平方和的分布。

这些分布的个数是 $\chi^2$ 分布变量的自由度 $\nu$，$\nu$ 是这个分布的一个参数。

$\chi^2$ 分布变量的概率密度函数是

$$f(\chi^2,\ \nu)=\frac{(\chi^2)^{(\nu/2)-1}}{2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)}\exp\left(-\frac{\chi^2}{2}\right)$$

$$\chi^2 \geqslant 0$$

14.t分布（司都顿分布）：独立随机变量的商的分布。其分子是标准正态变量，其分母是 $\chi^2$ 分布变量及其自由度的商的正平方根。

$\chi^2$ 的自由度数是 $t$分布变量的自由度数 $\nu$。

$t$分布变量的概率密度函数是

$$f(t,\ \nu)=\frac{1}{\sqrt{\pi\nu}}\frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}$$

15.$F$分布：两个独立 $\chi^2$ 分布变量的商的分布，每个变量都除以各自的自由度。$\chi^2$ 分布变量的自由度数分子为 $\nu_1$、分母为 $\nu_2$，$F$分布变量的自由度数也是这个顺序。

$F$分布变量的概率密度函数是

$$f(F,\ \nu_1,\ \nu_2)=\frac{\Gamma[(\nu_1+\nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}(\nu_1)^{\nu_1/2}(\nu_2)^{\nu_2/2}$$

$$\times\frac{F^{(\nu_1/2)-1}}{(\nu_1 F+\nu_2)^{(\nu_1+\nu_2)/2}}$$

$$F\geqslant 0$$

16.对数正态分布：可以取从 $a$ 到 $\infty$ 之间任一值的连续随机变量 $X$ 的概率分布，其概率密度为：

$$f(x)=\frac{1}{(x-a)\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\log_e(x-a)-\mu}{\sigma}\right)^2\right]$$

$$x>a$$

式中：$\mu$——$\log_e(x-a)$的平均值；

$\sigma$——$\log_e(x-a)$的标准差。

注 1：变量$\log_e(x-a)$的概率分布是正态分布；$\mu$和$\sigma$ 分别是此变量的数学期望和标准差。

注 2：常用$\log_{10}$代替$\log_e$。此时

$$f(x)=\frac{0.4343}{(x-a)\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\log_{10}(x-a)-\mu}{\sigma}\right)^2\right]$$

式中$\mu$和$\sigma$分别是$\log_{10}(x-a)$的平均值和标准差。

17.二项分布：离散随机变量 $X$ 的概率分布，如 $x\varepsilon$、0，1，2，……，$n$`，于是

$$P_r[X=x]=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$0<p<1$$

18.普瓦松分布：离散随机变量$X$的概率分布，如 ε 0，1，2，……}，于是

$$P_r[X=x]=e^{-m}\frac{m^x}{x!}$$

式中 $m$ 是正参数。

注：普瓦松分布的数学期望和方差都等于 $m$。

19.指数分布：可以取从0到$+\infty$之间任一值的连续随机变量 $X$ 的概率分布，其分布函数为：

$$F(x)=1-e^{-\lambda x}$$

$$\lambda>0$$

用$(x-a)$代替 $x$（但 $x\geqslant a$）可以使此概率分布普遍化。

20.单元：

1)一个实际的或传统的对象，对此可以进行一组观测，或

2)一规定数量的材料，对此可以进行一组观测，或

3)一个观测到的值，可以是定性的（特征），也可以是定量的（尺度）。

21.总体：所考虑的各个单元的总计（全体）。

一个总体的每一个清楚地规定的部分称做子体。

对于一个随机变量，其概率分布可用来规定该变量的总体。

22.特征：系一种性质，它帮助区别一规定总体的各个

单元。

区别可以是定量的(用变量),也可以是定性的(用特性)。

23.检验：为了衡量或区分一个特征所做的一种运算。

24.观测值：观测或试验结果确定的一个特征的值。

25.绝对差：两个值之间的差的绝对值。

26.极差：一个定量特征的最大观测值和最小观测值之间的差。

27.组：在定量特征的情况下,变化总区间区分成的连续区间中的每一个小区间。

28.组限：规定一个组的上限和下限的数值。

29.组中值：一个组的上限和下限的算术平均值。

30.组距：一个组的上限和下限之间的差。

31.绝对频数或容量：一个总体、一批、一个样本、一个组，等等的单元的数量。

32.累计绝对频数：在定量特征的情况下,其值小于或等于某一给定值的，或者小于或等于某一给定组上限的单元的数量。

33.相对频数：观测得的一特殊值（或落在一给定组的值）的次数与观测值总数的比值。

34.累计相对频数：在定量特征的情况下,其值小于或等于某一给定值的，或者小于或等于某一给定组的上限的单元相对频数。

35.频数分布：一个特征的值与其绝对或相对频数之间的关系。

36.直方图：一个连续变量的频数分布的图示法。

各组由直线比例坐标轴之一上的邻接区间表示。各组的绝对（或相对）频数由以这些区间为底的矩形表示，矩形面

积与这些组的绝对（或相对）频数成正比。

37.线条图：一个离散变量频数分布的图示法。与变量的绝对（或相对）频数成比例的直线段垂直地画在一轴上，在此轴上变量的值由直线比例表示。

38.累计绝对（相对）频数折线图：折线是由连接各点构成。这些点的横坐标是每个组的上限，纵坐标是累计绝对频数，或是累计相对频数。

39.算术平均值：各值的和被所有值的数量除所得的商。

40.（算术）加权平均值：每个值与其权（每一个值都有一个指定的非负数系数，称做权）的乘积的和除以权的总和所得的商。

41.中位数：如 $n$ 个数值按大小递增排列并标号 1 到 $n$ 如 $n$ 是奇数，则 $n$ 个值的中位数是 $\frac{n+1}{2}$。如 $n$ 是偶数，则中位数位于第 $\frac{n}{2}$ 个值与第 $\left(\frac{n}{2}+1\right)$ 个值之间。除非另有规定，这两个值的算术平均值可以取作中位数。

42.平均偏差：与一原点的各个偏差（所有偏差都是正号）的算术平均值。

注：虽然取中位数为原点时平均偏差最小，但通常所选的原点是算术平均值。

43.方差：一个分散性的量度，它以与算术平均值的偏差的平方的平均值为基础。随所考虑的情况而定，与算术平均值的偏差的平方和可以除以偏差的数量或除以偏差的数量减 1。

例如，对于一列 $n$ 个观测值 $x_1$，$x_2$，……，$x_n$，其平均值 $\bar{x}=\frac{1}{n}\Sigma x_i$，，可以用下式之一表示其方差：

$$\frac{1}{n}\Sigma(x_i - \bar{x})^2$$

或 $$\frac{1}{n-1}\Sigma(x_i - \bar{x})^2$$

重要的是要弄清楚所采用的定义。

注：第二个表达式常用$s^2$表示。

44.标准（偏）差：方差的正平方根。

45.偏差系数：标准差与算术平均值的绝对值的比值。此比值也可以用百分数表示。

46.协方差：1阶和1阶的中心矩。

$$\frac{1}{n}\sum_{ij} n_{ij}(x_i - \bar{x})(y_i - \bar{y})$$

式中：$n_{ij}$——配对变量$(x_i, y_i)$的观测频数；

$n = \sum_{ij} n_{ij}$——所观测配对变量的总数量。

47.相关系数：它表示两个变量线性相关的密切程度，在数值上它等于两个变量的协方差与它们标准差乘积的比值。

48.回归曲线：在两个变量$X$和$Y$的情况下，$Y$对$X$的回归（经验）曲线对变量$X$的每一个值$x$提供变量$Y$的一个中心值（即平均值）$y_{(x)}$（或$\bar{y}$）。

当回归曲线可以看作是一条直线时，这种回归称做“线性”回归。

在此情况下，$Y$对$X$的回归系数是回归直线方程$y = y(x)$中$x$的系数。

49.回归曲面：在三个变量$X$、$Y$和$Z$的情况下，对于变量$X$和$Y$的每一对值$(x, y)$，$Z$都有一个数学期望。当$x$和$y$的值变动时，$Z$的数学期望所形成的曲面称为“Z对