

主办单位

- 中国科协学会工作部
- 中国科协普及工作部
- 中国科协继续教育中心
- 中央电视台
- 中国质量管理协会
- 中国电子学会
- 中国仪器仪表学会
- 中国通信学会

史定华 编著

可靠性数学

人民邮电出版社

《可靠性工程与管理》电视讲座和函授班教材 三

主办单位

中国科协学会工作部	中国科协普及工作部
中国科协继续教育中心	中 央 电 视 台
中国质量管理协会	中 国 电 子 学 会
中国仪器仪表学会	中 国 通 信 学 会

可 靠 性 数 学

史定华 编著

人民邮电出版社

登记证号(京)143号

内 容 提 要

本书为《可靠性工程与管理》电视讲座和函授班教材之三。书中系统讲述可靠性数学的基本理论与方法,包括应用概率统计基础、可靠性的数量特征、估计与检验、系统可靠性、故障树分析以及可靠性的最优化问题。

本书适用于可靠性与质量管理人员、工程技术人员及大专院校数学工程与管理专业的师生学习参考。

可 靠 性 数 学

史定华 编著

*

人民邮电出版社出版

北京东长安街27号

北京广益印刷厂印刷

人民邮电出版社发行

*

开本: 787×1092 1/16 1989年4月第一版

印张: 13¹²/₁₆ 页数: 110 1993年1月北京第2次印刷

字数: 344千字 印数: 12 001—17 000册

ISBN 7-115-03810-4/Z·131

定 价: 9.50元

《可靠性工程与管理》电视讲座和函授班

主办单位

中国科协学会工作部 中国科协普及工作部
中国科协继续教育中心 中 央 电 视 台
中国质量管理协会 中国电子学会
中国仪器仪表学会 中国通信学会

教育委员会

顾问:

盛树仁 高镇宁 宋季文 刘 恕
张五球 叶柏林 陈保定 马怀祖

主任委员:

宋直元

副主任委员 (以姓氏笔划为序):

丁俐丽 牛田佳 邓震垠 (常务) 朱玉龙
成银生 李传卿 陆廷杰 罗国英 林振申
苑郑民 钟 良 (常务) 魏学兴

委员 (以姓氏笔划为序):

马 林 马桂夫 么子臣 王圣媛 王相龙
宁云鹤 史定华 汪元江 刘宗仁 过元柄
庄异君 陈 刚 陈章豹 牟致忠 何国伟
郎锋军 杨为民 周 济 周维田 杨定亚
林中强 单永铮 茆侍松 陆洪时 徐运忠
张庆龙 殷鹤林 程光辉 傅光民 裘履正

承办教学单位

上海电子学会可靠性与质量管理专业委员会
上海第二工业大学
中国电子产品可靠性与环境试验研究所
《电子技术》杂志社

教研组

组长、教育录像和教材主编：

傅光民

副组长：

裘履正

成员（以姓氏笔划为序）：

马怀祖	史定华	许 康	庄异君	牟致忠
罗 威	林中强	卓礼章	郁时霖	茆侍松
费鹤良	倪正铭	夏春镗		

前 言

可靠性 (Reliability), 是产品的重要质量指标。可靠性高, 意味着寿命长、故障少, 可靠性低, 意味着寿命短、故障多。电视机的平均无故障工作时间, 汽车的平均无故障行驶公里数, 运载火箭的发射成功率等都是产品的可靠性指标。

可靠性工程与管理是 40 年代以来迅速发展起来的新兴综合学科, 涉及数学、物理、化学、电子、机械、环境、管理以及人机工程等各个领域。它致力于研究提高产品可靠性, 包括从原材料、元器件、零部件到整机及系统的各个环节。从研究、设计、制造到使用及维修的全寿命周期, 是一个十分复杂的系统工程。国内外的实践表明, 可靠性工程与管理技术的应用, 为企业与社会带来了巨大的经济效益, 因而引起世界各国的普遍重视与关注, 纷纷投入大量人力物力进行研究和推广应用。产品的可靠性, 已经成为当今国际和国内市场竞争的焦点。

1987 年 9 月, 国家经委、国家机械委、国防科工委、劳动人事部、广播电影电视部、中国科协联合发文决定, 由中国科协、中央电视台、中国质协、中国电子学会、中国仪器仪表学会、中国通信学会联合主办全国性的可靠性工程与管理电视讲座和函授班。中国通信学会为牵头单位, 承担组织工作。上海电子学会可靠性与质量管理专业委员会、上海第二工业大学、中国电子产品可靠性与环境试验研究所、《电子技术》杂志社承办教学工作。聘请上海市第二工业大学可靠性研究室主任傅光民同志、上海自动化仪表研究所高级工程师裘履正同志负责组成教研组, 承担制订电视讲座及函授班的教学大纲, 提出课程设置及详细提纲, 组织编写课本及全套书面教材的工作。经教学双方共同努力, 第一期教学取得了良好成果。

近几年来, 产品可靠性工作开始得到重视和加强, 产品可靠性规划、设计、试验、失效分析、评审、鉴定、指标考核和相应的管理、监督逐步开展, 特别是国标、行(部)标和产品质量分等标准中对产品可靠性作为限期必须考核的项目实行以来, 机电产品质量有了一定程度的提高。但是, 发展极不平衡, 我国产品与工业先进国家的同类产品相比, 仍有较大差距。开展可靠性工作要从人才培养入手。为了进一步在全国范围培养大批可靠性工程技术人才和可靠性管理人才, 促进可靠性工作的全面开展, 大幅度提高我国产品可靠性质量, 1992 年 9 月, 人事部、中国科协、机械电子部、国防科工委、广播电影电视部、航空航天部、邮电部、国家技术监督局联合发文决定, 由中国科协学会部、普及部、继续教育中心, 中央电视台和上述四个全国性学会(协会)联合主办第二期可靠性工程与管理电视讲座和函授班。中国通信学会为牵头单位, 承担组织工作。为了加强组织领导, 聘请国家计委、国务院电子信息系统推广应用办公室、上述发文单位、主办单位和有关院校、科研院所、企业等单位的领导干部、专家、学者组成可靠性工程与管理电视讲座和函授班教育委员会(第二届)。教学承办单位、教研组负责人同上届。

联合发文指出: “提高产品质量, 是国民经济和社会发展的一项长期战略任务, 在加快改革开放和经济发展的新形势下, 尤为重要。提高产品可靠性是提高产品质量和提高产品社会、经济效益的基础, 也是繁荣市场, 促进出口, 保证产品上台阶, 在商品竞争中赢得主动

权的必要条件。”“各级经济管理部门，各有关部门，各企业和相关的科研、设计、生产、监督、试验、使用、维修等部门，应当把可靠性技术培训列为专业技术人员和管理人员在职教育和岗位培训的一个重要内容，要充分利用举办电视讲座和函授班的有利条件，结合实际情况组织本系统本地区有关人员积极报名参加学习。已经建立继续教育登记制度的单位，可将参加本次教学的学员考试成绩登记入册。并将这次培训作为考核审查可靠性工作开展情况的一个方面的依据。”

根据岗位培训的实际需要，教学分设管理班与工程班。管理班学员是有关企业、研究所以及主管部门的领导干部与管理人员。学员应收看中央电视台第一套节目播出的 22 集电视教学片（每集 50 分钟），自学《可靠性工程与管理电视讲座教材》、《可靠性管理》一书；工程班学员是工程技术人员及可靠性与质量管理工作人员。学员除收看电视讲座，学习《可靠性工程与管理电视讲座教材》、《可靠性数学》、《可靠性物理》、《可靠性管理》外，选学《可靠性设计》、《锡焊技术与可靠性》、《可靠性试验》、《环境试验》、《机械可靠性》等五门课程中的一门。《可靠性教学辅导教材》作为参考。

教学认真贯彻理论联系实际、学以致用方针，注意系统性、实用性，着重阐明物理概念，给出定性分析、定量计算方法及运用实例，避免繁琐的数学推导。内容以民用电子设备为重点，讲授可靠性通用技术，兼顾仪器、仪表、通信、航天、航空、轻工等系统的部分应用实例。通过电视讲座及函授学习，可以帮助学员了解可靠性工程与管理的发展历史与重要意义；掌握可靠性工程与管理的主要工作内容及本岗位的可靠性技术（可靠性设计、制造、试验及管理）；了解部分企业、研究所行之有效的实践经验；从而提高可靠性工程与管理技术水平，提高产品可靠性。为此，教材请国内有较丰富工程与教学实践经验的同志编写，总结国内外富有成效的可靠性工作案例，参考国内外可靠性书刊及论文，有较广泛的适用性和较高的实用性，可作为在职可靠性岗位培训的教材，也可作为大专院校可靠性与质量管理专业的参考教材。

在教材编写与出版以及电视教学片摄制过程中，得到中央有关部委、有关全国性学会（协会）、上海及各地工厂企业、研究所、大专院校、人民邮电出版社等 50 多个单位 200 多位同志的大力支持、指导和帮助，在此一并表示衷心的感谢。

由于时间紧迫，工作量很大，组织编写、摄制系统性的可靠性教材及电视教学片尚属首次，缺乏经验，不妥之处，敬请读者批评指正。

可靠性工程与管理电视讲座和函授班教育委员会

1992 年 9 月

编 者 的 话

本书是《可靠性工程与管理》电视讲座和函授班教材之五。主要是为从事电子产品设计人员进行可靠性设计方面的在职教育和岗位培训之用。

可靠性设计是可靠性工程的主要内容。产品的可靠性主要是由设计所决定，所以可靠性工作必须从设计抓起，设计人员必须补上可靠性设计这一课。

根据《可靠性工程与管理》电视讲座和函授班教育委员会的教学方针，本书以民用电子产品为重点，兼顾工业和国防电子设备与系统，贯彻理论联系实际，学以致用方针。全书着重讲授可靠性设计技术的应用，扼要讲授可靠性设计的理论知识。使读者通过对本书的学习，能基本掌握可靠性设计的一般方法。本书吸收了目前国内外出版的有关可靠性书刊文献、培训班讲义和可靠性标准中的部分内容，也有相当部分是作者多年来从事可靠性工程的现场实践、应用实例和可靠性教学工作的总结。

本书经倪正铭、周元初两位同志审核，并提出了许多宝贵的意见，最后经教研组长、教材主编傅光民同志审定，特此表示衷心感谢。

由于可靠性设计技术正处在发展之中，加之作者水平的限制，成书时间的仓促，谬误之处在所难免，祈请广大读者给予批评指正。

作 者

1988年2月于上海

目 录

第一章	绪 论	1
第二章	应用概率统计基础	3
第三章	可靠性中的数量指标	36
第四章	可靠性中的估计问题	63
第五章	可靠性中的检验问题	101
第六章	系统可靠性分析	124
第七章	故障树分析	156
第八章	可靠性中的最优化问题	175
附表A	泊松分布部分和 $\sum_{i=0}^m \frac{e^{-m} m^i}{i!} \leq 1 - \alpha$	205
附表B	标准正态分布函数 $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$ 的数值表 ($x \geq 0$)	206
附表C	χ^2 分布的上侧分位数 $\chi^2_{\alpha}(n)$ 表	207
附表D	F 分布的上侧分位数 $F_{\alpha}(n_1, n_2)$ 表	209

第一章 绪 论

可靠性工程学是一门涉及面广泛的综合性学科。目前它已渗透到自然科学以至社会科学的许多领域。

可靠性工程学由可靠性理论和技术组成。理论与技术之间通过可靠性规范沟通。可靠性理论包括定量分析理论和故障机理理论。其中,故障机理理论和各门技术学科关系较为密切。如电子元件的失效机理与半导体物理有关,机械零件的破损规则与材料强度理论以及断裂力学有关等等。当前,故障机理理论以可靠性物理给予概括。定量分析理论主要运用数学中的概率、统计、优化、代数等工具,对所有的产品都适用,共性较强,已经形成一门崭新的学科——可靠性数学理论。它的研究目标是定量评价产品的可靠性,并在一定的经济技术要求下,使产品进一步优化。

可靠性数学在可靠性工程学科中之所以重要,也是因为在定量研究产品可靠性当中,始终离不开概率论和数理统计。可靠性中的数量指标都是用概率论中的术语精确地描述的。产品的失效分布、产品的可靠性指标评定都是由寿命试验数据和现场统计数据利用数理统计方法得到的。除此之外,许多可靠性技术都要用概率、统计、优化、代数的工具才能表述清楚。可以毫不夸张地说,没有可靠性数学就不可能创立新的可靠性技术,也无法评价所采取手段的优劣。例如,一个航空发动机厂,每年要生产大批的发动机,并将其安装到各种类型的飞机上去。工厂希望利用现有的元器件尽量组装成可靠性高的发动机,在配套安装到各类飞机上去时,也希望可靠性尽量高。采取什么样的可靠性技术才能达到这一目标呢?经过可靠性数学理论家的深入研究,发现一种优化的组装技术可以办到。首先把元器件的可靠性分类,然后把最高可靠性的元器件装在一台发动机上,把次高可靠性元器件装在第二台发动机上……在配套安装到飞机上时,则需把最高可靠性的发动机和最低可靠性的发动机冗余,等等,就能达到目标。又如,在建设全国的微波中继通信系统时,各种可靠性的微波站应如何设置才能使整个系统的可靠性最高;如何缩短大批量生产的彩色电视机的平均寿命鉴定试验台时数,以及比较新生产工艺的产品寿命是否高于老生产工艺的寿命等等,这一切,没有可靠性数学理论作后盾是无法办到的。

可靠性数学理论研究的是机器维修问题。所用的技巧是根据爱尔兰(Erlang)等人研究电话占线问题而发展起来的排队论。早期人们试图把泊松过程作为电话呼唤的输入,从而为使用指数分布作为复杂设备的失效规律奠定了基础。

30年代后期,威布尔(Weibull)等人研究了材料的疲劳寿命和有关极值分布理论。以他的名字命名的威布尔分布是可靠性中最著名的分布之一,这一分布特别适用于描述机械产品可靠性。后来,龚伯尔(Gumbel)在他的著作中精心地提供了许多数据来说明如何用各种极值分布去渐近地表示产品的寿命分布。

50年代前期,可靠性理论中的一些领域,特别是寿命试验,电子产品与导弹可靠性问题开始受到数理统计学家与军工工程师们的热心关注,当时最著名的组织AGREE在1957年

发表的报告中论述了可靠性试验的要求、接受界限,以及储备在可靠性方面的作用。

1951年开始,爱泼斯坦(*Epstein*)等人开始研究寿命试验,这一研究标志着指数寿命分布在寿命试验中广泛应用的开始。在*AGREE*报告发表后,指数寿命分布的地位就更加稳固。

1956年,摩尔和仙农(*Moore & Shannon*)发表了一篇在数学上很有意义的论文。这是一篇涉及继电器网络可靠性的文章。该文受到了冯·诺依曼(*Von Neumann*)试图描述由复杂器官组成的人脑所具有的特殊功能及高可靠性的启发。

1958年,斯密思(*Smith*)很好地总结了更新理论中的结果,这套理论能用来解决替换问题。这一时期在用半马尔可夫过程解决系统的维修性问题方面也有所发展。

在新一代商用喷气机上遇到的振动问题的强力推动下,拜尔帮姆(*Birnbaum*)等人对动态载荷条件下的结构寿命提出了一个精采的统计模型。

60年代初期,三根久、拜尔帮姆等人独立地提出了单调关联系统的理论。接着有关寿命分布类的研究非常活跃,这在下面要提到的巴罗和普罗斯钦的书中有很完整的叙述。与此同时,一个与统计相依有关的重要概念是随机变量间的牵连性(*association*)。牵连性概念虽然受到单调关联结构思想的启发,但它有更深刻的含义。为了让工程师们接受这一概念,马歇尔与奥肯(*Marshall & Olkin*)引进了多变量(指数)分布概念。

70年代可靠性工程研究重点转向了故障树分析。其部分原因是受到人们对核电站安全性的关注所推动。许多有深远影响的结果发表在巴罗等人编辑的《可靠性和故障树分析》文集中。这一时期,维修策略最优化的研究也很活跃。

如果说本世纪70年代是故障树的天下,那么80年代是网络可靠性的时代。这主要是由于计算机及其网络可靠性的重要性所致。伯克利的一批博士研究生将制约数(*domination*)的概念引入了网络可靠性计算。这标志着开始运用复杂的图论技术研究系统可靠性计算。大型复杂系统可靠性的计算复杂性问题已开始受到人们的重视。

致力于可靠性理论研究的统计学家中最有意义的动向是,对贝叶斯(*Bayes*)方法用于归纳推理的有效性的认识愈来愈深刻。

在可靠性最优化问题的研究中,组合最优化方法也已开始得到应用。

可靠性理论和排队论的联系正在受到我国学者的重视。

美国在1965年由巴罗(*Barlow*)和普罗斯钦(*Proschan*)出版了第一本《可靠性数学理论》专著。1975年以《可靠性和寿命试验的统计理论》的书名进行了扩充和修改。1981年又再版了一次。苏联也在1965年由格涅坚科,别里亚夫,索络维也夫出版了第一本名为《可靠性理论中的数学方法》的专著。1983年又出版了《可靠性数学理论问题》专著,系统地总结了苏联学者的成果。日本在1982年出版了三根久和河合一著的《可靠性、维修性数学》。我国则在1983年出版了戴树森等编著的《可靠性试验及其统计分析》,在1984年出版了茆诗松和王玲玲编著的《可靠性统计》,在1986年出版了曹晋华和程侃编著的《可靠性数学引论》。

第二章 应用概率统计基础

可靠性数学以产品的时间质量为其研究对象。产品的性能指标一般都可用仪器加以测量，而产品的时间质量——可靠性指标该如何度量呢？常用的办法是统计现场失效数据或在实验室模拟现场使用条件（也可以加速）进行产品寿命试验。这样我们就可以得到一批数据。但这些数据往往具有很大的随机性和分散性。造成随机性和分散性的原因是多方面的，有各种各样的偶然因素，也有某些必然规律。为了定量地给出可靠性指标，就必须正确分析这些数据，研究数据随机波动的规律。这就是为什么概率统计——一门研究随机现象、处理随机数据的学科，在可靠性数学中占有着中心的位置。

本章将扼要介绍应用概率统计的基本概念、定理、公式、例子等等，为读者学习后面各章做准备。

2.1 统计资料的整理

“母体”、“个体”和“子样”是数理统计学中最常用的名词。母体也称“总体”，它指的是我们研究对象的全体。而个体指的是母体中的一个基本单元。例如，我们要研究某种型号电视机的寿命，那么所有这种型号电视机的寿命就是一个母体，而其中每一台电视机的寿命则为一个个体。显然，母体的性质是由每一个个体的性质所确定的。但若把母体中所有个体都一一加以研究，则会遇到两种主要困难：首先，在一般情况下，母体包含的个体为数众多，因此不可能对所有个体都进行研究；其次，在可靠性问题中，由于个体寿命的测定是具有破坏性的，因此从经济的观点看，不能对所有个体都做这种破坏性的试验。

由于以上两种原因，为了推测母体的性质，常从母体中抽取一部分个体来加以研究。这些被随机地抽中的个体，形成一个“子样”或称为“样本”。子样所包含个体的数目，称作“子样大小”或“样本容量”。子样中的个体则称为样品。

对于子样通过试验或实测所得到的一组数据，通常没有一定的规律和次序，因此必须经过整理，才能显示出子样数据的规律性。所以，数据的整理工作是进行统计分析的第一步。

下面叙述通过子样观察值（数据）来建立频数（频率）直方图的方法。

假定子样容量为 n 并得到 n 个观测数据，分别以 $x_1, x_2, \dots, x_i, \dots, x_n$ 表示，建立直方图的步骤是：

（1）找出观测数据的最大值和最小值，求出两者之差（也称极差）：

$$R = \max_{1 \leq i \leq n} \{x_i\} - \min_{1 \leq i \leq n} \{x_i\} \quad (2.1)$$

（2）根据子样大小进行分组，通常可分成5~10组，也可采用斯特奇斯（Sturges）建议的经验公式来确定分组的数目 K ，即

$$K = 1 + 3.32 \lg n \quad (2.2)$$

(3) 根据组数 K 和极差 R 决定组距 C 。如果按等距间隔分组, 则

$$C = \frac{R}{K-1} \quad (2.3)$$

(4) 确定各组的端点值, 端点值通常比原始数据的精度高一位, 使得原始数据不落在组区间界限上。初始端点值 x_0 按下式计算:

$$x_0 = \min\{x_i\} - \frac{D}{2}$$

式中 D 或为组距 C , 或为测量单位;

(5) 算出各组的频数 V_i ;

(6) 计算各组的频率

$$f_i = \frac{V_i}{n}$$

(7) 以端点值为横坐标, 以频数 V_i 或频率 f_i 为纵坐标, 画出直方图。前者称为频数直方图, 后者称为频率直方图。一般都采用频率直方图。在画频率直方图时, 纵坐标的刻度要使得单位面积刚好等于频率的 $1/100$, 以保证频率分布直方图的总面积等于1。

【例1】某型号电子计算机在60年代统计的60个寿命数据(单位小时)如下:

91, 77, 55, 58, 105, 32, 64, 64, 64, 46, 96, 47, 67, 30, 84, 53, 78, 37, 97, 25, 60, 80, 73, 47, 40, 69, 50, 77, 13, 79, 43, 82, 61, 37, 73, 38, 49, 48, 50, 66, 76, 57, 44, 60, 87, 29, 112, 87, 43, 21, 57, 59, 22, 64, 75, 47, 71, 69, 55, 72。

初看, 看不出这组数据的性质。但经过排列, 便可发现小于30的数值很少, 大于100的更少, 而40~70左右的最多。然而更详细的情况就不清楚了。为了更好地了解这组数据, 我们来画出它的直方图, 步骤如下:

(1) 这组数据的最大值为112, 最小值为13, 两者之差

$$R = 112 - 13 = 99$$

(2) 确定分组数 K

$$K = 1 + 3.3 \lg 60 = 6.868$$

取 $K = 7$;

(3) 按等距离分组, 其组距为

$$C = \frac{99}{7-1} = 16.5$$

(4) 确定初始端点值

$$x_0 = \min\{x_i\} - \frac{C}{2} = 4.75$$

(5) 进行分组统计, 数出频数, 算出频率, 就可得到如表2-1所示的频数(率)表。

(6) 以表2-1中端点值为横坐标, 频数 V_i 或频率 f_i 为纵坐标, 可分别画出频数或频率直方图, 见图2-1。

【例2】某种电子设备共18台, 从开始使用到发生失效的时间数据(单位天), 按从小到大排列如下:

16, 29, 50, 68, 100, 130, 140, 190, 220, 270, 280, 340, 410, 450, 520, 620,

800, 1100。

表 2—1

例 1 的频数 (率) 表

分组端点值	频 数 V_i	频率 $f_i = \frac{V_i}{n}$	组 中 值
4.75~21.25	2	0.0333	13
21.25~37.75	7	0.1167	29.5
37.75~54.25	14	0.2333	46
54.25~70.75	17	0.2833	62.5
70.75~87.25	15	0.2500	79
87.25~103.75	3	0.0500	95.5
103.75~120.25	2	0.0333	112

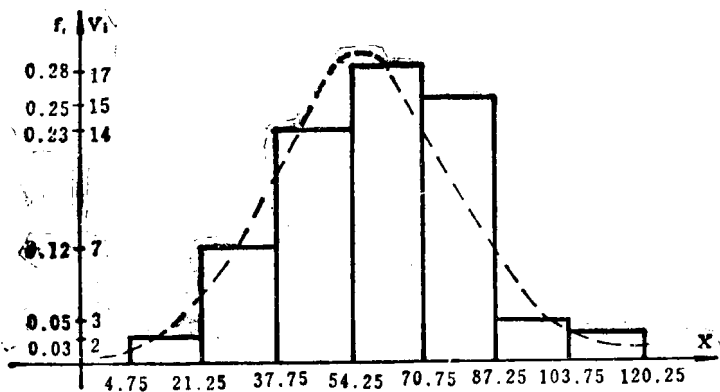


图 2—1 60个寿命数据的直方图

这组数据的特点初看是大部分失效数据都密集在前面。但为了更好地了解这组数据，仍需将数据进行分组统计，作出频率直方图。按一般步骤得：

- (1) 极差 $R = 1100 - 16 = 1084$;
- (2) 估算 $K = 5$ ，我们选取 $K = 6$;
- (3) 求出 $C = 216.8$ ，我们选取 $C = 220$;
- (4) 初始端点值取

$$x_0 = \min\{x_i\} - \frac{1}{2} = 16 - \frac{1}{2} = 15.5$$

- (5) 进行分组统计，得频数 (率) 表如表 2—2 所示；

表 2—2

例 2 的频数 (率) 表

分组端点值	频 数 V_i	频率 $f_i = \frac{V_i}{n}$	组 中 值
15.5~235.5	9	0.5	125.5
235.5~455.5	5	0.2778	345.5
455.5~675.5	2	0.1111	565.5
675.5~895.5	1	0.0556	785.5
895.5~1115.5	1	0.0556	1005.5

- (6) 画直方图如图 2—2。

比较上述两个频率直方图，不难看出，这两组数据是具有不同特点的。例 1 的数据大多数都密集在这组数据的中心，具有中间大、两头小的分布特性；例 2 的数据大多数都密集在初始阶段，具有前面较密、后面稀疏的分布特性。在进行数据统计分析时，常常需要寻求一条光滑曲线来拟合频率直方图的外形。如图 2—1 和图 2—2 中的虚线就是拟合出来的光滑曲线，称作“实验频率曲线”。尽管所研究数据的直方图各不相同，但它们的实验频率曲线都具有下列共同的特征：

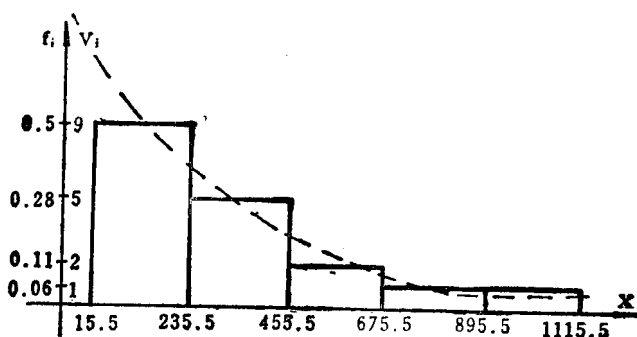


图 2—2 18 个失效数据的直方图

(1) 曲线的纵坐标恒为非负值；

(2) 曲线与横坐标轴所包围的面积都等于 1。

图 2—1 的实验频率曲线呈钟型，后面将会知道，它可用解析式

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

来表示，称为正态密度函数。

图 2—2 的实验频率曲线呈滑波下降型，它用解析式

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

来表示，称为指数密度函数。

2.2 观察数据的特征值

一组观察数据 $x_1, x_2, \dots, x_n$ ，除了将它画成频数（率）直方图找出它的统计规律外，还常常用几个能反映它的主要特征的数值来描述。本节介绍的均值和中值反映了数据的集中位置；而方差、标准差、极差则反映了数据的分散程度。

2.2.1 均值和中值

均值或平均值，又称平均数或算术平均数。设给定一组数据 $x_1, x_2, \dots, x_n$ ，则

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.4)$$

称为数据 $x_1, x_2, \dots, x_n$ 的均值。

为什么要求一组数据的均值呢？因为一组数据参差不齐，杂乱无章，能用一个有代表性的数来反映它们，既醒目又能为人们所接受。例如，某种铬镍钛合金制成的 5 个标准试件，在疲劳试验机上测得的以 kc （千循环数）为单位的疲劳寿命为

535, 604, 376, 901, 565

从这5个数据很难给人深刻的印象,但其平均寿命

$$\frac{1}{5}(535+604+376+901+565)=\frac{2981}{5}=596.2kc$$

却简单明了地反映了这种试件耐疲劳的情况。

下面我们来分析用均值作为一组数据代表的合理性。对给定的一组数据 $x_1, x_2, \dots, x_n$,原则上可以用任何一个数 c 去“代表”它们,问题只在于代表得好不好。 $x_i - c$ 反映了 c 偏离 x_i 的程度,也就是 c “代表” x_i 的好坏。为了消除符号的影响,用

$$(x_1 - c)^2 + (x_2 - c)^2 + \dots + (x_n - c)^2 = \sum_{i=1}^n (x_i - c)^2$$

来衡量 c “代表” $x_1, x_2, \dots, x_n$ 的好坏。显然,代表性最好的 c 应使 $\sum_{i=1}^n (x_i - c)^2$ 达到最小。

定理1 均值 $\bar{x}$ 是唯一使 $\sum_{i=1}^n (x_i - c)^2$ 达到最小的数 c 。

证: (1) $\sum_{i=1}^n (x_i - \bar{x}) = 0$

因为

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$n\bar{x} = \sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

$$= (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) + n\bar{x}$$

所以

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

(2) $\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2$

因为

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2$$

$$= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - c) \sum_{i=1}^n (x_i - \bar{x}) + n(\bar{x} - c)^2$$

由(1),上式中间一项为零,所以

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2$$

(3) 只有 $c = \bar{x}$ 才使 $\sum_{i=1}^n (x_i - c)^2$ 达到最小

由于平方数总大于等于零, 从 (2) 可得

$$\sum_{i=1}^n (x_i - c)^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$$

这说明最小值是 $\sum_{i=1}^n (x_i - \bar{x})^2$, 仅当 $c = \bar{x}$ 时才达到。

在求 $x_1, x_2, \dots, x_n$ 的均值时, 如果 x 中有相同的, 自然可以合并, 不妨设不同的只有 k 个值 $a_1, a_2, \dots, a_k$, 并且 a_i 出现了 n_i 次, $i=1, 2, \dots, k$ 。此时

$$x_1 + x_2 + \dots + x_n = n_1 a_1 + n_2 a_2 + \dots + n_k a_k$$

因此

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^k n_i a_i = \sum_{i=1}^k a_i \cdot \frac{n_i}{n} \quad (2.5)$$

上式说明均值 $\bar{x}$ 与 a_i 在全部数据中所占比例有关。比值 $\frac{n_i}{n}$ 越大, $\bar{x}$ 受 a_i 的影响就大; $\frac{n_i}{n}$ 越小, $\bar{x}$ 受 a_i 的影响就小。

把式 (2.5) 加以抽象, 就产生了加权平均数的概念。设给了一组数据 $x_1, x_2, \dots, x_n$, 又给了一组正数 $p_1, p_2, \dots, p_n$, 且 $\sum_{i=1}^n p_i = 1$, 则

$$\bar{x} = p_1 x_1 + p_2 x_2 + \dots + p_n x_n = \sum_{i=1}^n p_i x_i \quad (2.6)$$

称为 $x_1, x_2, \dots, x_n$ 的加权平均数, $p_1, p_2, \dots, p_n$ 称为相应的权。很明显, 当 $p_1 = p_2 = \dots = p_n = 1/n$ 时, 即一视同仁, 加权平均数就是均值。“权” p_i 反映了数据 x_i 在平均时的重视程度; 怎样合理地决定“权”, 是一个重要的问题。

从 $\bar{x}$ 的定义可以看到, $x_i (i=1, 2, \dots, n)$ 个别的很大的值或很小的值, 对 $\bar{x}$ 的影响较大。有时这些极端值并不反映真实的情况, 则可除掉数据中的一个最大值和一个最小值再平均。但更简便的是采用中值。

中值, 又称中位数。将数据 $x_1, x_2, \dots, x_n$ 按大小次序排列成 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ 。若 $n = 2k + 1$, 则中值 $M_c = x_{(k+1)}$; 若 $n = 2k$, 则中值 $M_c = \frac{x_{(k)} + x_{(k+1)}}{2}$ 。

例如前面铬镍钛合金标准试件的 5 个疲劳寿命数据按大小次序排成: $x_{(1)} = 376, x_{(2)} = 535, x_{(3)} = 565, x_{(4)} = 604, x_{(5)} = 901$, 故 $M_c = x_{(3)} = 565$ 。

2.2.2 方差、标准差、极差

方差, 意即平均的平方偏差。对给定的一组数据 $x_1, x_2, \dots, x_n$ 来说, $\bar{x}$ 是与这 n 个数最接近的数。 $x_i - \bar{x}$ 是 x_i 偏离 $\bar{x}$ 的程度, $(x_i - \bar{x})^2$ 是 x_i 与 $\bar{x}$ 偏差的平方, 这 n 个数据与 $\bar{x}$ 的平均的偏差平方就是