# 统计学简史

**H.O.Lancaster著**

**中国人民大学统计学院 吴喜之 译**

**1. 起源，分布**

**2. KarPearson时代，1890－1920**

**3. R.A.FISHER时代，1921-1936**

**4. NEYMAN－PEARSON时代，1937－1949**

**5. 现代，1950—**

  简写sb.代表有某人,具体的人物在文章后面列出。不在文章中列出,是因为人数太多,列在文章中，显示不出统计学的主要人物.还难以记忆.所以单列出来.

## 1. 起源，分布

统计是初产生于对国家的研究，特别是对其经济以及人口的描述。当时现代数学尚未形成。因此那时的统计史基本上是经济史的范畴。现代统计主要起源于研究总体（population），变差（variation）和简化数据（reduction of data）。第一个经典文献属于John Graunt（1620－1674），其具有技巧的分析指出了把一些庞杂、令人糊涂的数据化简为几个说明问题的表格的价值。他注意到在非瘟疫时期，一个大城市每年死亡数有统计规律，而且出生儿的性别比为1.08，即每生13个女孩就有14个男孩。大城市的死亡率比农村地区要高。在考虑了已知原因的死亡及不知死亡年龄的情况下，Graunt估计出了六岁之前儿童的死亡率，并相当合理地估计出了母亲的死亡率为1.5％。因此，他从杂乱无章的材料中得出了重要的结论。他还给出了一个新的生命表。

sb1.利用了Breslau的记有死亡年龄的数据，改进了Graunt的生命表并引进了死亡率的定义。瑞士数学家sb2.提出了平稳生命表的概念。sb3. 等人最早讨论退休金和人寿保险的方案。Thomas Robert Malthus（马尔萨斯）（1766－1834），sb4.  ，Ronald Aylmer Fisher（费歇）（1890－1962），及 William Feller（费勒）（1906－1970）等人用渐趋复杂的数学来研究生命表的理论，这对人类及其它总体的动方学描述具有显著意义。

William Petty（1623－1687）是Graunt同时代的经济学家及朋友。他认为需要建立中央统计部来利用人口统计学的知识；由行政区利用列出记录年龄，性别，婚姻状况等细节的记录表格来收集数据；要有出生，死亡，婚姻，收入，教育和商业等方面的统计数据。

当时在研究诸如死亡等时间序列时，Graunt注意到了随机的起伏；但他仅以机械的术语加以描述一把这些与钟表运动的忽动忽停相联系。实际上，这种不规则的变化也影响赌博和天文学。因此，其后进一步导致了随机误差的误差分布概念的出现。

赌博产生了第一个机会事件的模型：如果硬币就骰子的每一面都有相同概率，则导致估计抛一个均衡的硬币所出现的正面次数或掷一个均衡的骰子的总点数。更一般地，Abraham de Moivre（棣美佛）（1667-1754）导出了对二项分布的一个近似；这使每一个概率都等于正态曲线下的一块面积，这是一种的中心极限定理。sb5. 导出了对男子出生比例的类似的渐近公式。sb6.  以弱大数定律支持了对大样本均值的使用。sb7. 计算了同分布随机变量和的精确分布，同样也支持了对大样本均值的使用。

在天文学中，要对一些运动星体位置的未知参数进行估计，通常某种意义上“最好的”估计都是来源于一些注定不和谐的观察值，因为只要观察值在数量上超过参数，就会产生度量误差。sb8. ，Thomas Bayes（贝叶斯）（1702－1761），sb9. ，sb10. ，sb11. ，sb12.  和sb13. 都在研究这个问题。后来被Friedrich Gauss（高斯）（1777－1855）解决。sb14. 用统计方法证明了双星的存在。

然而，认定现代统计理论是由精算科学，人口学和天文学的需要而发展来的观点是不正确的；事实上，它是由心理学，医学，人体测量学，遗传学和农业的需要发展出来的。

直到1830年，几乎所有的经验分布都是关于一维误差或一个非数值变量。在1830年之后，天文学家和社会学家 sb15.使得诸如身高体重之类的度量值的变量的经验分布通俗化。他在生物统计研究中大量利用了理论二项分布和正态分布。后来sb16. 报告了在普鲁士兵团中由马踢造成的受伤事故，发现Poisson（普阿松）分布和官方统计学有关。在计算血红细胞数目上，Poisson分布也被sb17. 所用。从那时起，该分布被大量地用于计数的试验中，比如闪光的计数。

在生物学上，统计方法使得Johann Gregor Mendel（孟德尔）（1822－1884）认识到某些主要遗传基因的存在，它们在0，l和2三个水平显现，其中水平0（双隐性）能和水平1和2区别开来。他能确定有相同或不相同的水平的个体之间交配的结果，而且提出了某些生物学事件等价于掷一个硬币的模型；他能对任意交配的结果给出概率并用实验来验证其假设。

虽然经济学没有产生超出用初等理论来求解问题，但在较早的医学统计中却夹生了有意思的问题。sb18.  和s19. 开始了建立疾病分类的困难课题；这些工作人员保存了精确和完整的所有病例的记录，并且能给出和预后有关的统计数字。sb19. 能有利用跟踪调查的方法反驳以当时广泛滥用的放血疗法。他的三个学生是值得一提的：sb20.写了一本医学统计的教科书；书中有应用Simeon Dents Poisson（1780－1840）理论来对两个比例进行检验的许多应用；sb21. 和他的不知名的数学顾问对一系列分娩热病例给出了有趣的分析，证明该病是传染的，这优于任何十九世纪的类似研究；sb22. 在官方统计学中建立了新的惯例。

更直接的原动力来自于遗传学（确切地说是优生学）。Francis Galton（1822一1911）在1886年研究了两代豌豆重量之间的相关时发现了Y关于一个正态变量X的线性回归及类似于椭圆的等概率线。从此，多元正态分布就经常出现在文献之中；而两个和三个变量的正态分布在Lapface时就已经知道了。该联合分布能够由互相独立的正态随机变量的线性变换而构造，例如Gb23. 和sb24. 所做. sb24. 2在最小二乘理论上导出了一般形式的多元正态分布；sb25.把xTAx化简为平方和并确定了该常数值。这些人都未对A-1＝V的非对角线元素感兴趣，这里V是协方差矩阵。Galton后来说“这些误差或偏差正是我想要研究并了解的。”

正态分布在理论统计中扮演了一个非常重要角色。有许多理由来说明这一点；一般来说，如果一个模型包含着正态分布的几个非平凡特性，则它必须具备所有的特性。

在 1895年，Karl Pearson（皮尔森）（1857－1936）认识到更理论的统计分布的需要，并且得到作为微分方程（Pearson方程组）解的密度函数；和另外一些统计学家一样，Andrei andreevic Markov（马尔科夫）（1856－1922）不愿意用Pearson分布方程组，因为即使得了皮尔森Г曲线作为一个极限分布，也没有明显的模型来产生它们。Markov进一步证明Pearsonχ2统计量为样本尺寸乘以sb26.的离散系数。sb27. 利用取独立第二项变量和的方法得到二项变量的联合分布。许多作者，比如sb28. 已经参与了发展该思想；但是许多其它思想已经被用来获得联合分布。在KarlPearson的方法不能产生更多的联合分布之后，sb29.认为一个更具有生产价值的方法可能存在于随机过程的领域中。

## 2. KarPearson时代，1890－1920

到1920年为止的英国生物统计学派的主要就为（i）收集并化简了许多经验数据；（ii）定义了具有多重和总相关系数ρ的联合正态分布，还定义了估计误差的联合分布；（iii）关于拟合度的χ2检验，比较观察分布和理论分布，包括由sb30. Herbert Edward Soper（1865－1930）引进的条件Poisson变量；（iV）分析列联表，特别是利用χ2统计量；（V）当边缘分布充分细分时由最大似然法估计ρ；（Vi）当边缘分布没有充分定义时估计ρ；（Vii）由一个统一的参数估计系统来描述一组曲线，即矩方法；（Viii）利用正态定理到遗传选择问题；（iX）通往独立性一般定理的某些进展；（X）通往估计和检验估计精确性的一个理论的进展；（Xi）构造了适当的表。

与此同时，在法国的 Feli院sb31. ，sb32. 及sb33. 和在俄国的sb34. ，sb35. 和sb36.作出许多贡献，特别是把数据的数学处理严格化。

## 3. R.A.FISHER时代，1921-1936

所有的皮尔森的方法都可以应用于大样本，而且可以对方差作出较精确的估计。但对于出现在实际应用中的小样本，这些方法就未必奏效了。sb37.国此导出了一个检验；按照R.A.Fisher的建议该检验在作了一个变换之后成为现在熟和的t-检验。

Fisher以其四篇值得纪念的论文开创了一个新纪元；相关系数估计的精确分布；协调一致了Mendelian和生物统计对遗传学的不同方法；正确解释了列联表；估计和推断的一般定理。在1920年之后，在Rothamsted实验室，Fisher发展了有广泛应用价值的方差分析和试验与分析的理论。Fisher有很强的数学功底，特别是在组合论（combinatorics）方面，他能吸引其他数学家作为助手。他对应用领域的选择是很幸运的；研究结果都能立即应用并有明显的经济效应；能够有效地简化假设，比如误差的正态性和独立性；和一些顽固的教条斗争；试验的花费都很低；没有伦理问题。许多重要的步骤用来发展上面提到的Pearson学派的工作的一些分支。在（iii）和（iV）中的许多重要问题被解决了；给出了正确的自由度；

K.Pearson已经为该目标前进了一段；在（Vii）中Fisher发明了更有效的方法来估计；他拒绝了用矩方法来确定分布；在(Vi)中Fisher和Yate发表了统计表。

Fisher在误差分布和互相独立性的假设使其能用正交变换来保持线性和二次型之间的独立性，这样就可合理地利用t-检验和F-检验。Fisher看到农业试验能利用更复杂的设计。于是，双向表的影响能够按地理因素（行和列）及处理来分别分析。这能推广到n维Latin方的应用，把处理用于Latin方相应的不同字母的位置上；该方法通过实行Graeco－Latin方来实现。Fisher及其助手和同事研究了设计问题，缺损值问题，非正交性等等；这些人包括sb38.，sb39. ，他们后来在美国特别有影响。此外，还有 sb40.，sb41. ，sb42. ，及sb43.  。

## 4. NEYMAN－PEARSON时代，1937－1949

Jerzy Neyman（1894－1981）及 Egon sharpe Pearson（1895一1980）在一系列的杰出的文章中澄清了推断理论，特别是有关显著性检验的基本原理一其合理性以往是常被批评。早期的显著性检验为关于二项变量之间或均值之间的，它们被 K.Pearson推广至χ2检验，被R.A.Fisher推广到F-检验，推广了Student T-检验。Neyman和 E.S.Fearson看出，为了更有效，应该考虑与待检验的零假设相对应的备选假设。他们在这样的检验中设立两种误差并因素导致了他们的基本引理，似然比检验，及势的概念；他们顺便验证了大多数常见的显著性检验的应用；他们还引进了置信限；但是他们的体系从未被Fisher所承认。Neyman和Pearson的工作影响了许多人，特别是美国人。

## 5. 现代，1950—

统计变得越来越数学化了。为了解对分布和推断理论的一般描述，需要测度论；Fourie分析成为研究波动最自然的工具；在分析方差的推断上，和在具对称性的设计以及在诸如 Graeco Latin方及 Steiner三元的特别结构的代数的推断上需要应用群论和数论。组合理论能用于编码理论和有限几何。因此统计数学成为纯粹数学的一部分，并且因其在各种领域的广泛应用而被研究。因为通常的统计检验已经彻底地研究了，而且往往被置身于某些具体应用领域的实际工作者所应用，所以在研究人员和实际工作者之间出现了一个距离；但是这种现象在其他开拓性的领域中也能看到。

电子计算机已经带来了巨大的变化。数据，比如海洋学中水面的高程，电磁能（特别是无线电波）的流量，工业过程的状态，生物的状态，都能用计算机收集；没有计算机这些是不实际或不可能的。计算机节省了大量人力，特别是在同样输出的重复计算上，例如在计算多元分析的相关系数和其它检验统计量时。由于计算软件包可用于所有通常的检验，特别是关于方差分析，则节省更多。高速计算使得有可能运用匹配和排列检验。当分布不能写成一个封闭的分析公式时，显著性水平也能计算；另外，计算机能用 Monte Carlo方法计算每一个事件的概率或近似显著性水平。由于利用软件包很方便，有时导致统计问题欠考虑而产生的结论，特别是在多重比较上。

模型在统计和科学工作中的作用现已被广泛承认；虽然基于应用领域的经验和知识，模型的选择在某种程度上是美学上的和任意的；但是一旦模型被选定，所有的推断都是数学的，用不着进一步的假设或原则。

所用的推断体系在某种程度上也是任意的；备选体系已经被大量研究。基于信仰的推断不再扮演重要的角色。贝叶斯模型在 Fisher时代曾一度失色，之后又被更广泛地应用。信息论已被引进；多数统计推断看来仍然以和Neyman—Pearson理论一致的方式来运作，运用在K.Pearson和Fisher时代引人的检验。

现代国家的增长的能力和兴趣要求以低花费收集更多的数据。sb44. 有远见地建议概率抽样应补充到人口普查方法中。这样的抽样已经在其被sb45. P引进之后成为在印度和其它地方的标准实践。被sb6. 所推广的工业质量控制方法也有类似的意图。

许多新的分支或专门化和应用已经被发展了：决策论，时间序列，多元分析，经济计量学，博奕论，临床试验，非参数推断，序贯分析，数学生物分类学，及可靠性。数量统计及其应用正在继续发展和扩大。

## 著名人物

sb1. Edmond Helley（哈雷）（1656－1742）  
sb2.  Leonhard Euler（欧拉）（1707－1783）  
sb3.  John DeWitt(625－1672)  
sb4.  Alfred James lotke（1880－1949）  
sb5. Pierre simon Laplace（拉普拉斯）（1749－1827）  
sb6.  Jacob Bernoulli（伯努利）（1654－1705）  
sb7. Tomas simpson（辛普森）（1710－1761）  
sb8. Rog.er Cotes（1682－1716）  
sb9. Euler  
sb10. Johann Toblas Mayer（1723－1762）  
sb11. Rudger Josif Boskovic（1711－1787）  
sb12.  Laplace  
sb13. Adrien Marie Legendre（勒让德）（1752－1833）  
sb14. John Michell（米歇尔）(1724－1793）  
sb15. Adolphe Jacques Quetele(1796－1874)  
sb16. Ladislaus von Bortkiewicz（1868－1931）  
sb17. Ernst Abbe（1840－1905）  
sb18. Philippe Pinel（1745-1826）  
sb19. Pierre Charles alexandre Louis（1787－1872）  
sb20. Jules Gavarret（1808－1890）  
sb21. Oliver Wendell Holmes（1809－1894）  
sb22. William Farr（1807－1883）  
sb23. iovanni Antonio amedeo Plana（1781－1863）  
sb24. Irenee－Iules Bravais（1811－1863）  
sb25. Auguste Bravais（l820－l884）  
sb26. Wihelm Hector Bichard albrecht Lexis（1837一1914）  
sb27. Walter Frank Raphael Weldon（1860－1906）  
sb28. Alexander Craig Aitken（1895－1967）  
sb29. Sergei Natanovic Bernstein（1880－1968）  
sb30. Herbert Edward Soper（1865－1930）  
sb31. x Edouard Justin emile borel（1871－1956），  
sb32. Maurice Frenchet（1878-1973）  
sb33. Jules Henri Poincare（1854－1912）  
sb34. Alekandr Aleksandrovic Cuprov（1874－1926）  
sb35. A.A.Markov  
sb36. Vsevolod Ivanovic Romanovsky（l879－1954）  
sb37. William Sealy Gosset（1876-1937）  
sb38. Maurice Stevenson Bartlett（1910－）  
sb39. William Gemmell cochran（1909－1982）  
sb40. David John Finney（1917－）  
sb41. Joseph Oscar Irwin（1898—1982）  
sb42. Kanneth Mather（1911一）  
sb43.  Frank Yates（1902－）  
sb44. Antlers Nicolai Kiaer（1838－1919）  
sb45. Prasanta chandra Mahalanobis（1893－1972）  
sb6. Andrew Shewhart(1891－1967）