

现代管理中的数理统计方法

张维群 编著

人民邮电出版社

国 际 市 场 学

——打入国际市场的战略

(美) 富兰克林 R·鲁特 著
万安培 陈桂芳 屈定坤 译
邵培德 戴炳源

上海社会科学院出版社

责任编辑 高品瑚

封面设计 邹越非

国际市场学

——打入国际市场的战略

〔美〕富兰克林·鲁特 著

万安培 陈桂芳 屈定坤 邬培德 戴炳源 译

上海社会科学院出版社出版

（上海淮海中路622弄7号）

新华书店上海发行所发行 上海中行印刷厂常熟分厂印刷

开本787×1092 1/32 印张 9.625 字数 213千字

1991年3月第1版 1991年3月第1次印刷

印数1—3300

ISBN7—80515—557—7/F·171

定价：3.30元

引 言

数理统计方法是研究随机现象的一种数学方法。众所周知，随机现象广泛地存在于社会实践中，而它常常可以用一个或几个随机变量来表述。例如，某一个工厂生产电子管，由于受各种随机因素的影响，各个电子管的寿命不尽相同。因此，该工厂生产的电子管的寿命便可视为一个随机变量。对于该工厂的质量管理人员而言所关心的问题往往是此随机变量的某些统计特性如何。如该工厂某日所生产的电子管寿命长短的分布如何？其平均寿命如何？是否达到了所规定的标准？影响电子管平均寿命的因素有哪些？等等。而我们在研究这些问题时又不能将生产的所有电子管测量一遍，一种可行的办法是从所生产的电子管中抽取一部分测其寿命，然后对所获得的数据进行整理、分析，对所关心的问题作出相应的推断。

数理统计方法就是以概率论为基础来研究如何用有效的方法收集、整理和分析所获得的局部数据，从而对所考察的整体问题作出相应推断。于是，数理统计方法通常可分为两方面：一是如何有效地收集数据，即抽样方法；另一是如何由收集到的数据合理地进行整理、分析，对所研究的问题作出推断，即统计推断。本书主要介绍后者，但需指出的是，抽样方法与统计推断两者之间通常有着相辅相承的关系，不同的抽样方法有着不同的统计推断方法，要得到更有效的推断结果，也往往需要与相应的抽样方法配合。

目前，数理统计方法已被广泛地应用于各个领域内，特别是在现代企业管理中，统计方法已成为制定各种决策的有力工具。

目 录

引 言	1
第一章 抽样分布	1
§ 1 总体、样本、统计量	1
一、总体	1
二、样本	1
三、统计量	4
§ 2 数理统计中的几个常用分布	6
一、 χ^2 分布	6
二、t 分布	10
三、F 分布	14
四、Wishart 分布	17
五、Wilks 分布	18
§ 3 正态总体下的抽样分布	19
第二章 点估计	27
§ 1 点估计的标准	28
一、无偏估计	28
二、最小方差无偏估计	31
三、有效估计	32
四、相合估计与渐近正态估计	37

§ 2 最大似然估计.....	38
一、最大似然估计的基本思想原理	38
二、似然方程	40
(一) 单参数情形 (二) 多参数情形 (三) 多元正态总 体期望与协差阵的估计	
三、最大似然估计的性质	48
(一) 不变性 (二) 有效性 (三) 相合性 (四) 渐近正 态性	
§ 3 矩估计.....	50
一、矩估计法的基本思想原理	50
二、矩法方程	54
第三章 区间估计	56
§ 1 置信区间.....	56
§ 2 相对似然函数法.....	58
§ 3 正态总体期望与方差的区间估计.....	68
一、置信区间的建立	68
二、常用的置信区间	76
三、应用举例	76
§ 4 置信域.....	80
第四章 参数假设检验	83
§ 1 假设检验的基本概念.....	84
一、原假设与备选假设	84
二、拒绝域与接受域	84
三、两类错误	85
四、功效函数	86

§ 2	广义似然比检验·····	88
§ 3	正态总体期望与方差的假设检验·····	95
	一、检验法的建立·····	95
	二、常用结果·····	103
	三、应用举例·····	103
	四、方差不等时的双样本问题·····	108
	(一) 大样本情况 (二) 成对样本情况 (三) 方差成比例情况 (四) Scheffé 方法	
§ 4	二维正态总体相关系数的假设检验·····	113
§ 5	多元正态总体期望与协差阵的假设检验·····	120
	一、协差阵已知期望的检验·····	120
	二、协差阵未知期望的检验·····	122
	三、两个总体期望的检验·····	126
	四、协差阵的检验·····	129
第五章	非参数假设检验·····	132
§ 1	顺序统计量·····	132
§ 2	位置参数的假设检验·····	134
	一、单样本问题·····	134
	(一) 符号检验 (二) Wilcoxon 符号秩检验	
	二、成对双样本问题·····	142
	(一) 符号检验 (二) Wilcoxon 符号秩检验	
	三、独立双样本问题·····	145
	(一) 中位数检验 (二) Wilcoxon 秩和检验 (三) 游程检验	
	四、多样本问题·····	154
§ 3	分布的检验·····	157

一、正态性检验	157
(一) Shapiro-Wilk 检验 (二) 偏度、峰度检验	
二、拟合度检验	163
(一) Pearson χ^2 检验 (二) Колмогоров 检验	
三、齐一性检验	175
(一) Pearson χ^2 检验 (二) Смирнов 检验	
§ 4 独立性检验	181
一、联立表检验	181
二、Spearman 检验	184
三、Kendall 检验	188
§ 5 随机性检验	191
一、Daniel 检验	191
二、游程检验	193
第六章 线性回归分析的基本方法	195
§ 1 线性回归的数学模型	195
§ 2 线性回归中的参数估计	199
一、 β 的最小二乘估计	199
二、 $\hat{\beta}$ 的性质	201
三、 σ^2 的估计	202
四、有关 $\hat{\beta}$ 和 $\hat{\sigma}^2$ 的分布	203
§ 3 线性回归中的假设检验	206
一、带线性约束条件的最小二乘估计	206
二、有关参数 β 的假设检验	211
(一) 线性回归方程的显著性检验 (二) 回归因子的 显著性检验	
§ 4 线性回归中的区间估计	218

一、关于回归系数的区间估计·····	218
二、预测·····	221
三、实例·····	222
§ 5 一元线性回归·····	227
一、散点图·····	227
二、回归系数 β_0 , β_1 及方差 σ^2 的估计·····	228
三、关于回归直线斜率与截距的假设检验·····	230
四、斜率与截距的置信区间·····	231
五、预测区间·····	232
§ 6 多个因变量的线性回归·····	235
一、 β 和 Σ 的估计·····	236
二、假设检验·····	238
第七章 回归变量的选择·····	242
§ 1 回归变量多少对回归效果的影响·····	242
§ 2 逐步回归·····	247
一、逐步回归的数学模型·····	247
二、 T_i 变换·····	250
三、逐步回归的计算公式与步骤·····	254
§ 3 最优回归模型的评选准则·····	264
一、PRESS 统计量·····	265
二、 C_p 统计量·····	268
第八章 影响回归效果的原因分析及改进措施·····	276
§ 1 异方差性·····	276
一、异方差性的检验——残差图法·····	277
二、改进措施——加权最小二乘估计·····	278

§ 2 序列相关性	286
一、序列相关性的检验— <i>D.W</i> 检验法	286
二、改进措施——变换法	291
§ 3 多重共线性	296
一、多重共线性的检验—— <i>VIF</i> 法	298
二、改进措施——岭回归	302
§ 4 异常值	312
一、检查异常值的意义	312
二、检查异常值的统计方法	313
第九章 曲线回归	320
§ 1 变换法	320
一、可化为线性回归的曲线回归类型	320
二、应当注意的问题	332
§ 2 正交多项式	335
一、时间序列模型与多项式回归	335
二、正交多项式	336
三、变量的还原	344
第十章 方差分析	348
§ 1 单因素方差分析	348
一、数学模型	348
二、统计分析	351
(一) 假设检验 (二) 参数估计 (三) 多重比较	
三、等重复试验的方差分析	364
§ 2 双因素方差分析	366
一、数学模型	366

二、统计分析	369
(一) 假设检验 (二) 参数估计 (三) 多重比较	
三、无重复试验的双因素方差分析	386
§ 3 三因素方差分析	393
§ 4 多元方差分析	401
第十一章 判别与聚类	408
§ 1 判别分析	408
一、距离判别	408
(一) Mahalanobis 距离 (二) 判别准则	
二、Bayes 判别	416
三、Fisher 判别	422
(一) 建立判别函数的准则 (二) 判别函数的求法	
(三) 判别准则	
§ 2 聚类分析	429
一、最短距离法	430
二、最长距离法	436
三、重心法	438
四、类平均法	442
五、离差平方和法 (Ward 法)	443
六、其他方法	447
第十二章 多元相关分析	455
§ 1 主成分分析	455
一、主成分分析的基本原理	455
二、累积贡献率	460
三、基于相关阵的主成分分析	463

§ 2 因子分析	466
一、因子分析的统计模型	466
二、载荷矩阵的确定	469
三、因子正交旋转	474
§ 3 典型相关分析	480
一、引言	480
二、典型相关系数与典型相关变量	481
三、典型相关系数的估计和检验	488

附表	493
----------	-----

附表 1 标准正态分布函数值表	493
附表 2 t 分布上侧分位数表	495
附表 3 χ^2 分布上侧分位数表	497
附表 4 F 分布上侧分位数表	501
附表 5 二项分布函数值表	513
附表 6 $M(m, v_0, k)$ 表	516
附表 7 相关系数检验的临界值表	518
附表 8 偏度、峰度检验临界曲线	519
附表 9 柯尔莫哥洛夫检验的临界值表	521
附表 10 斯米尔诺夫检验的临界值 α 表	523
附表 11 符号检验表	525
附表 12 Wilcoxon 符号秩检验的临界值表	526
附表 13 两样本秩和检验的临界值表	529
附表 14 游程总数检验表	530
附表 15 Spearman 检验临界值表	532
附表 16 Kendall 检验临界值表	533
附表 17 $D.W$ 检验临界值表	534
附表 18 Shapiro-Wilk 检验统计量 W^2 系数 $a_i(W)$ 表	

.....	536
附表 19 Shapiro-Wilk 检验临界值表	538
附表 20 t 化极差上侧 α 分位数表	539
附表 21 R -学生检验临界值表	542
附表 22 正交多项式表	546
参考文献.....	555

第一章

抽 样 分 布

§ 1 总体、样本、统计量

一、总体

所研究对象的全体称为总体。组成总体的每个元素称为个体。然而在实际问题中，我们关心的往往不是组成总体的每个个体，而是它们的某个或某些数量指标的统计特性如何。例如，一批电子管为一个总体，而每一个电子管便为此总体中的一个个体，我们关心的通常并不是这一个个电子管，而是它们某个数量指标或某些数量指标如何，如电子管的寿命的统计特性如何。用概率论观点看，这样的某一个或一些数量指标可以视为定义在总体上的一维或多维随机变量。为了简便，以后我们便将它们称为总体，分别记为 X （一元总体）和 X （多元总体）。

二、样本

对总体 X 进行 n 次观测，可得到 n 个观测值 $x_1, \dots, x_n$ ，可用 x_i ($i=1, \dots, n$) 表示第 i 次观测所得到的数据。我们称 $(x_1, \dots, x_n)'$ 为来自总体 X 的样本值。

由于对总体抽样观测是随机的, 所以样本值 $(x_1, \dots, x_n)'$ 可视为 n 维随机向量 $(X_1, \dots, X_n)'$ 的一个可能取值, 我们称此随机向量为来自总体 X 的样本。

n 称为样本容量。

在不变的条件下, 对总体 X 进行 n 次独立重复观测得到容量为 n 的样本 $(X_1, \dots, X_n)'$, 则称此样本为简单随机样本。简单随机样本显然应具有以下两个特点:

- (1) $X_1, \dots, X_n$ 与总体 X 同分布;
- (2) $X_1, \dots, X_n$ 相互独立。

由于简单随机样本具有上述两个特点, 因此, 当总体 X 的分布函数为 $F(x)$ 时, 则简单随机样本 $(X_1, \dots, X_n)'$ 的分布函数为

$$F_n(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i) \quad (1.1.1)$$

如果总体 X 为离散型的, 其概率分布为

$$P\{X=x_k\} = p_k, \quad k=1, 2, \dots$$

则简单随机样本 $(X_1, \dots, X_n)'$ 的概率分布为

$$P\{X_1=x_{k_1}, \dots, X_n=x_{k_n}\} = \prod_{i=1}^n p_{k_i} \quad (1.1.2)$$

如果总体 X 为连续型, 其分布密度为 $f(x)$, 则简单随机样本 $(X_1, \dots, X_n)'$ 的分布密度为

$$f_n(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i) \quad (1.1.3)$$

例 1.1.1 设总体 $X \sim B(1, p)$, 其中 $0 < p < 1$, $(X_1, \dots, X_n)'$ 为来自此总体的简单随机样本, 试求 $(X_1, \dots, X_n)'$ 的概率分布

解: 由于 X 的概率分布可表为

$$P\{X=x\} = p^x (1-p)^{1-x}, \quad x=0, 1$$

于是简单随机样本 $(X_1, \dots, X_n)'$ 的概率分布便为

$$\begin{aligned} P\{X_1=x_1, \dots, X_n=x_n\} &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}, \quad x_i=0, 1; i=1, \dots, n \end{aligned}$$

例 1.1.2 设总体 $X \sim N(\mu, \sigma^2)$, 其中 $-\infty < \mu < +\infty$, $\sigma^2 > 0$, $(X_1, \dots, X_n)'$ 为来自此总体的简单随机样本, 试求它的分布密度

解: 由于总体服从正态分布 $N(\mu, \sigma^2)$, 即其分布密度为

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

所以简单随机样本 $(X_1, \dots, X_n)'$ 的分布密度为

$$\begin{aligned} f_n(x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2\right\}. \end{aligned}$$

对于多元总体, 其样本、简单随机样本的概念与上述一元总体情况相同, 需要注意的是样本中每个分量是一随机向量。

例 1.1.3 设有 m 元总体 $X \sim N_m(\mu, \Sigma)$,

其中 $\mu = (\mu_1, \dots, \mu_m)'$ 为期望向量, $\Sigma = (\sigma_{ij})_{m \times m} > 0$ 为协差阵, 又设 $X_i = (X_{i1}, X_{i2}, \dots, X_{im})'$, $(i=1, 2, \dots, n)$, 为来自 X 的简单随机样本的第 i 个分量, 试求它的分布密度。

解: 由 $X \sim N_m(\mu, \Sigma)$, 即其分布密度为

$$\begin{aligned} f(x) &= f(x_1, \dots, x_m) \\ &= (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu)\right\} \end{aligned}$$

其中 $x = (x_1, \dots, x_m)'$ 所以简单随机样本 $(X_1, \dots, X_n)'$ 的联合分布密度便为