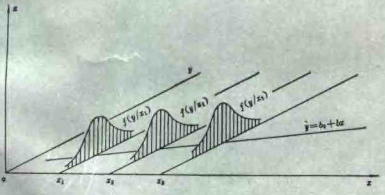
工商管理数理统计

宋念慈 编著



上海交通大学出版社

内 容 提 要

本书根据国家教委1990年颁布的经济数学基础教学大纲中的数理统计部分的内容提要设计和编写的，比较系统地介绍数理统计的基本概念、原理和方法。内容有：抽样分布、点估计、区间估计、假设检验、方差分析、回归分析。在抽样分布中强调了回置抽样与非回置抽样的差别，在假设检验中增加了多个正态总体方差齐性的检验，为方差分析的前提提供了检验方法。

本书配有大量应用性的例题和习题，对于计算题，在书末的习题答案中都给出了简要的解题过程，便于教师参考和学生学习。

本书可作为大学经贸专业本科的教材，也可作为经济贸易和企业管理人员的参考书。

责任编辑：冯 愈
封面设计：

(沪)新登字 205 号　　工商管理数理统计

出版：上海交通大学出版社
（上海市华山路1954号　邮政编码：200030）
发行：新华书店上海发行所　印刷：常熟市印刷二厂
开本：850×1168(毫米)1/32　印数：1—5000　字数：284000
版次：1994年4月　第1版　印次：1994年4月　第1次
印张：11　科目：311-208

ISBN 7-313-01297-7/O·212　定 价：9.30 元

导　言

随着改革开放的深入到社会各方面，工商企业所处的环境趋向于复杂多变，工商企业的管理工作者在决策时所遇到的不确定性因素也越来越多，以前那种单凭经验作决策的方式，已不能适应改革开放形势发展的要求。为了在工商企业管理工作中能作出预期效益较高而风险又小的决策，作为工商企业的管理工作者，面对大量的不确定性因素，要善于收集数据，会正确地进行统计分析，并对分析结果作出恰当解释，这就要求工商企业的管理工作者学习和掌握数理统计知识。

把数理统计的概念和方法，用来解决工商企业经营管理工作中的问题，这在美国、日本等发达国家已有数十年的历史，而且目前已渗透到工商企业经营管理中的很多环节，如生产计划的制订、成本的核算、生产流程的设计、工作人员业务水平的考核、产品的销售、市场需求量的预测等。

有一些与经济问题，工商管理问题有密切关系的学科，如运筹学、计量经济学、系统工程等，都离不开数理统计的概念和方法。

根据以上所述，可以说数理统计方法是工商企业管理人员必备的知识。

本书内容紧密结合经济贸易和工商管理方面的实际，在文理相互渗透上作了探索和尝试。

本书总结了编著者近20年的数理统计教学工作中的经验，基本概念叙述清晰，内容丰富，由浅入深，循序渐进，适于自学。

本书的姐妹篇——《概率论》，在1989年12月由对外经济贸易教育出版社出版。

由于著者水平所限，错误之处在所难免，恳请使用者批评指正。

编　著　者

1993年2月

目　录

导言
1　抽样分布 …… 1
1.1　总体与样本 …… 1
习题 …… 4
1.2　统计量及其分布 …… 4
习题 …… 24
2　点估计 …… 28
2.1　估计量的求法 …… 28
习题 …… 35
2.2　估计量的好坏标准 …… 35
习题 …… 39
3　区间估计 …… 41
3.1　总体均值的置信区间 …… 42
习题 …… 47
3.2　两个总体均值之差的置信区间 …… 48
习题 …… 57
3.3　总体比例的置信区间 …… 59
习题 …… 63
3.4　样本容量的确定 …… 64
习题 …… 68
3.5　正态总体方差的置信区间 …… 69
习题 …… 71
3.6　两个正态总体方差之比的置信区间 …… 72
习题 …… 75
4　假设检验 …… 77

4.1 基本原理和步骤 …… 77
习题 …… 84
4.2 检验总体的均值 …… 84
习题 …… 91
4.3 检验两个总体均值之差 …… 92
习题 …… 102
4.4 检验总体的比例 …… 103
习题 …… 106
4.5 检验两个总体比例之差 …… 107
习题 …… 110
4.6 检验正态总体的方差 …… 111
习题 …… 112
4.7 检验两个正态总体的方差 …… 113
习题 …… 117
4.8 检验多个正态总体的方差 …… 118
习题 …… 120
4.9 卡方分布在假设检验中的应用 …… 122
习题 …… 136
5 方差分析 …… 13
5.1 一个因素的方差分析 …… 139
习题 …… 15
5.2 两个因素的方差分析 …… 152
习题 …… 159
5.3 两个因素有交互作用时的方差分析 …… 161
习题 …… 163
6 回归分析 …… 171
6.1 一元线性回归中的参数估计 …… 171
习题 …… 180
6.2 回归直线方程的显著性 …… 182
习题 …… 187

6.3 利用回归直线进行预测 ……188
习题……192
6.4 一元非线性回归 ……193
习题……202
6.5 多元线性回归中的参数估计 ……203
习题……214
6.6 多元线性回归的第二种形式 ……215
习题……220
6.7 多元线性回归方程的显著性 ……221
习题……228
6.8 利用多元回归方程进行预测 ……230
习题……235
习题解答……238
附表 1 标准正态分布函数 $\Phi_{0,1}(x)$数值表 ……316
附表 2 t 分布上侧分位数 $t_\alpha(n)$表 ……321
附表 3 χ^2 分布上侧分位数 $\chi^2_\alpha(n)$ 表 ……325
附表 4 F 分布上侧分位数 $F_\alpha(a, b)$表……332
附表 5 相关系数临界值表……342
参考书目……344

1 抽样分布

本章主要介绍数理统计中的一些基本术语和基本概念，如总体、样本、抽样、统计量等以及几个重要的统计量的分布——抽样分布。

1.1 总体与样本

一、总体及其分布

在数理统计 (mathematical statistics) 中将所研究对象的全体称为总体 (population)，总体中每一个成员称为个体 (individual)。

例1.1.1 有一批茶具共100套，我们感兴趣的是这批茶具的质量，每套茶具可区分为一等品、二等品和等外品。这100套茶具的质量等级构成一个总体，每套茶具的质量等级就是个体。

例1.1.2 有一批尼龙布共1500匹，我们考察的是每一匹尼龙布上疵点个数，这1500匹尼龙布中每一匹布上疵点个数的全体构成一个总体，而每一匹尼龙布上各自的疵点个数就是个体。

例1.1.3 有一批20W节能灯泡共200个，对于它们的质量若只考察灯泡的寿命，那末这一批灯泡中每个灯泡的寿命的全体构成一个总体，每个灯泡的各自的寿命就是个体。

从上述例子中可看出，总体中的成员常常不是指成员本身，而是指成员的某种数量指标。在例1.1.2中，总体中成员是指每一匹尼龙布上的疵点个数。在例1.1.3中，总体中成员是每个灯泡的寿命。在例1.1.1中，如一等品用"1"表示，二等品用"2"表示，等外品用"3"表示，总体中成员是每套茶具的数量指标，它可以看成数"1"，"2"，"3"的集合。从这些例子还可以看出，数量指标取同一值的成员可以有几个，也就是每个值可以重复。总体是一个可以重复的数的集合。在例1.1.1的100套茶具中，一等品有56套，二

等品有32套，等外品有12套，因此在总体中，“1”占56/100，“2”占32/100，“3”占12/100。这说明在总体中各种数量指标是有一定比例（proportion）的，而取得某一规定指标是有一定可能性（possibility）的。这样总体与随机变量（random variable）联系起来了，在数理统计中说总体的分布（population distribution）就是指相应的随机变量的分布。为方便起见，总体的数量指标 ξ 有时简称总体 ξ，总体 ξ 的概率分布与数字特征采用的记号与随机变量的相应记号完全一样。

二、样本及其分布

从总体中取得的一部分个体称为样本，取得样本（sample）的过程称为抽样（sampling），样本中个体的个数称为样本容量（sample size）。表示样本时通常用小括号括起来，如在例1.1.2的总体中抽取一个容量为6的样本（3，4，0，5，8，7）。

在数理统计中，采用的抽样方法是随机抽样法，即样本中每一个个体是从总体中随意地抽取出来的。随机抽样分回置抽样（sampling with replacement）和非回置抽样（sampling without replacement）两种。如采用回置抽样，则总体中每一个个体在每次抽选中都有可能被选中。举例说，假如有一个会计想分析一批买主在一年内的购货情况，他决定从装有买主记录卡的卡片匣中随意取几个记录卡作为样本。若进行的是回置抽样，则从卡片匣中抽选一张记录卡，摘录所需的信息后，把这张卡片放回卡片匣，再抽选下一张记录卡，若以前抽出过的卡片再次被抽到，就再次摘录这张卡片上的信息，这样对同一个买主的信息，可以在样本中出现若干次；若进行的是非回置抽样，则抽出的卡片摘录其信息后，不再放回，再抽下一张，这样关于某个指定买主的信息，在样本中最多只能出现一次。实践中所采用的几乎都是非回置抽样。

从总体 ξ 中随机抽样得到的样本可以用 n 维随机变量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 表示。现在考察样本的概率分布，在回置抽样中，由于每取出一个个体后要放回、再取下一个，总体成分不变（分布不变），所以 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立同分布的，并且每一个随机变量 ξ_i 的分布

与总体ξ的分布相同；对于非回置抽样，要分两种情况：在总体中个体数是有限多个(有限总体)，因为每取出一个个体不放回，改变了总体的成分，所以随机变量$\xi_1,\xi_2,\cdots,\xi_n$不相互独立；在总体中个体数是无限多个(无限总体)，每取出的一个个体不放回，并不改变总体的成分，所以样本中的随机变量仍然是独立同分布的，并且每一个随机变量ξ_i的分布都相同于总体ξ的分布。

在实践中，有时遇到的是有限总体，而采用的是非回置抽样，此时，若样本容量n相对于总体容量N来说很小，则常常将非回置抽样看成回置抽样。实用上当$n/N\leqslant 0.05$时，大多数从事实际工作的统计工作者都把对有限总体进行非回置抽样，看做是回置抽样，这样样本中的$\xi_1,\xi_2,\cdots,\xi_n$看做是独立同分布的随机变量，且每一个ξ_i的分布都相同于总体的分布。

如果样本$(\xi_1,\xi_2,\cdots,\xi_n)$中各个$\xi_i$独立同分布，且每一个$\xi_i$的分布都相同于总体的分布，则称这个样本为简单随机样本(simple random sample)，这样的样本在数学上比较容易处理。

样本$(\xi_1,\xi_2,\cdots,\xi_n)$是$n$维随机变量，这是对具体进行一次抽样而言，在抽样后获得它的一组观察值$(x_1,x_2,\cdots,x_n)$称为样本值(sample value)。

设总体ξ的分布函数(distribution function)是$F(x)$，$-\infty<x<+\infty$，则简单随机样本$(\xi_1,\xi_2,\cdots,\xi_n)$的联合分布函数

$$F_n(x_1,x_2,\cdots,x_n)=F(x_1)\cdot F(x_2)\cdots F(x_n)$$
$$(-\infty<x_i<+\infty,i=1,2,\cdots,n)。$$

当总体ξ为离散型随机变量(discrete random variable)，设其概率函数为$f(x)=P\{\xi=x\}$，$x\in G$，则简单随机样本$(\xi_1,\xi_2,\cdots,\xi_n)$的联合概率函数

$$\begin{aligned}f_n(x_1,x_2,\cdots,x_n)&=P\{\xi_1=x_1,\xi_2=x_2,\cdots,\xi_n=x_n\}\\&=P\{\xi_1=x_1\}\cdot P\{\xi_2=x_2\}\cdots P\{\xi_n=x_n\}\\&=f(x_1)\cdot f(x_2)\cdots f(x_n)\\&\quad(x_i\in G,i=1,2,\cdots,n)。\end{aligned}$$

当总体ξ为连续型(continuous)随机变量，设其概率密度函数

为 $f(x)$，$-\infty<x<+\infty$，则简单随机样本$(\xi_1,\xi_2,\cdots,\xi_n)$的联合概率密度函数

$$f_n(x_1,x_2,\cdots,x_n)=f(x_1)\cdot f(x_2)\cdots f(x_n)$$
$$(-\infty<x_i<+\infty,i=1,2,\cdots,n)。$$

1.1 习　题

1.1.1　为什么可以将总体看成一个随机变量？

1.1.2　回置抽样与非回置抽样的区别在哪里？对无限总体来说两者的差别为何没有了？

1.1.3　在什么情况下，对有限总体进行非回置抽样可以看成是简单随机抽样？

1.1.4　设(ξ_1,ξ_2,ξ_3)为取自总体$\xi\sim N(\mu,\sigma^2)$的一个简单随机样本，写出随机变量(ξ_1,ξ_2,ξ_3)的联合概率密度函数。

1.1.5　设$(\xi_1,\xi_2,\xi_3,\xi_4)$为取自总体$\xi\sim P(\lambda)$的一个样本，写出$(\xi_1,\xi_2,\xi_3,\xi_4)$的联合概率函数。

1.2 统计量及其分布

样本是总体的代表及反映，但在取得样本之后，并不直接利用样本的 n 个观察值进行推断，而需要对这些值进行一番加工、提炼，把样本中所包含的有关我们感兴趣的信息集中起来，从而解决问题，这在数理统计中往往通过构造一个合适的依赖于样本的函数——统计量(statistic) 来达到的。统计量是样本的函数，且要求它不包含任何未知参数，因此它也是一个随机变量。统计量的概率分布称为抽样分布(sampling distribution)。下面介绍几个常用的统计量。

一、样本均值

设$(\xi_1,\xi_2,\cdots,\xi_n)$是来自某总体的一个容量为 n 的样本，称统计量

$$\overline{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i \qquad (1.2.1)$$

为样本均值(sample average)。

样本均值是一个随机变量，我们感兴趣的是它服从什么分布？它的均值和方差是什么？

样本均值服从什么分布，取决于两个因素：① 原来总体的分布；② 样本容量的大小。当原来总体服从正态分布时，不论样本容量的大小如何，样本均值都服从正态分布，既使当样本容量为1时，也是如此。但是当原来总体服从非正态分布时，这就要看样本容量的大小了。当样本容量 n 很大（大样本）时，根据中心极限定理（centrial limit theorem），样本均值近似服从正态分布(normal distribution)。当样本容量 n 不大(小样本)时，则样本均值不服从正态分布，不能按正态分布去作有关推断。样本均值的分布与原来总体的分布及样本容量大小的关系如表1.2.1。

表 1.2.1

原来总体的分布	样本大小	样本均值的分布
正态分布	大样本	正态分布
	小样本	正态分布
非正态分布	大样本	近似正态分布
	小样本	非正态分布

样本均值的分布类型确定了，那末它的均值和方差是什么呢？

设总体 ξ 的均值 $E(\xi)=\mu$，方差 $D(\xi)=\sigma^2$ 已知，从总体中抽取容量为 n 的样本，样本均值 $\overline{\xi}$ 的均值 $E(\overline{\xi})$ 与方差 $D(\overline{\xi})$ 的数值，也取决于两个因素：①总体 ξ 是有限总体(finite population)还是无限总体(infinite population)；②抽样方式是回置抽样还是非回置抽样。其中样本均值 $\overline{\xi}$ 的均值 $E(\overline{\xi})$ 比较简单，不论是

有限总体还是无限总体，是回置抽样还是非回置抽样，仍等于总体均值．既 $E(\overline{\xi})=\mu$。样本均值 $\overline{\xi}$ 的方差 $D(\overline{\xi})$ 则与总体的容量 N 及抽样方式有关，一般说，从无限总体抽样时，两种方式的抽样都可以看成是独立地抽样，所以 $D(\overline{\xi})=\sigma^2/n$。当从有限总体进行非回置抽样时，$D(\overline{\xi})=(\sigma^2/n)(N-n)/(N-1)$[4]。以上关系归纳于表1.2.2。

表 1.2.2

原来总体	抽样方式	$E(\overline{\xi})$	$D(\overline{\xi})$
无限总体 (μ,σ^2)	回置抽样	μ	σ^2/n
	非回置抽样	μ	σ^2/n
有限总体 (μ,σ^2)	回置抽样	μ	σ^2/n
	非回置抽样	μ	$(\sigma^2/n)(N-n)/(N-1)$

从表 1.2.2 看到，当在有限总体进行非回置抽样时，样本均值 $\overline{\xi}$ 的方差要比回置抽样时多出一因子 $(N-n)/(N-1)$。由于 $(N-n)/(N-1)\leqslant 1$，所以 $(\sigma^2/n)\geqslant(\sigma^2/n)(N-n)/(N-1)$，这说明：回置抽样时样本均值 $\overline{\xi}$ 的方差要大于非回置抽样时样本均值 $\overline{\xi}$ 的方差。因子 $(N-n)/(N-1)$ 称为有限总体修正系数。(finite population correction coefficient)。同时也可以看到，在无限总体中进行抽样时，由于

$$\lim_{N\to+\infty}\frac{N-n}{N-1}=1$$

所以在无限总体中进行非回置抽样时，样本均值 $\overline{\xi}$ 的方差可以用 σ^2/n 代替。

对于有限总体情形，当样本容量 n 相对于总体容量 N 来说很小时，有限总体修正系数可以不予考虑，因为当总体容量 N 远大

于样本容量 n 时，σ^2/n 和 $(\sigma^2/n)(N-n)/(N-1)$ 之间的差别是微不足道的。假如从容量为 10000 的总体中抽取一容量为 25 的样本，这时有限总体修正系数等于 $(10000-25)/9999=0.997599\cdots$，0.997599 与 σ^2/n 之积和 1 与 σ^2/n 之积几乎是相等的。当 $(n/N)<0.05$ 时，大多数统计工作者都不用有限总体修正系数。

例 1.2.1 某液化气罐的破裂压强 ξ(kg/cm²)近似服从正态分布 $N(2800,9216)$。从250个这类液化气罐中抽选一个容量为10的随机样本，并对每一个液化气罐作加压试验，直到它们都破裂为止。问样本中的液化气罐破裂的平均压强 $\overline{\xi}$ 不超过 2750kg/cm² 的概率有多大？

解 虽然是非回置抽样，但总体容量 N 相对于样本容量 n 来说很大，$n/N=10/250=0.04<0.05$。所以在 $D(\overline{\xi})$ 中的有限总体修正系数可以不予考虑，这样，样本均值 $\overline{\xi}\sim N(2800,9216/10)$，于是所求概率

$$P\{\overline{\xi}\leqslant 2750\}=P\left\{\frac{\overline{\xi}-2800}{\sqrt{9216/10}}\leqslant\frac{2750-2800}{\sqrt{9216/10}}\right\}$$

$$=P\left\{\frac{\overline{\xi}-2800}{\sqrt{9216/10}}\leqslant -1.647\right\}$$

$$=F_{0,1}(-1.647)=0.0495$$

例1.2.2 设某种切削工具其平均使用时间 $\mu=41.5$ h，标准差 $\sigma=2.5$ h，从仓库中随机抽取容量为 50 的样本。试估计这50个切削工具平均使用时间 $\overline{\xi}$ 在 40.5～42 h 之间的概率。

解 虽然未知总体分布类型，但样本容量 $n=50$，比较大，所以样本均值 $\overline{\xi}$ 近似服从正态分布。

若仓库中这种切削工具很多(超过 1000)，这时 $\overline{\xi}$ 近似服从正态分布 $N(41.5,2.5^2/50)$，所求概率

$$P\{40.5<\bar{\xi}<42\}=P\left\{\frac{40.5-41.5}{2.5/\sqrt{50}}<\frac{\bar{\xi}-41.5}{2.5/\sqrt{50}}<\frac{42-41.5}{2.5/\sqrt{50}}\right\}$$

$$=P\left\{-2.83<\frac{\bar{\xi}-41.5}{2.5/\sqrt{50}}<1.41\right\}$$

$$=F_{0,1}(1.41)-F_{0,1}(-2.83)$$

$$=0.9207-0.0023=0.9184。$$

若仓库中这种切削工具总数 $N=800$ 个，这时 $\bar{\xi}$ 服从正态分布 $N\left(\mu,\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}\right)=N(41.5,0.1173)$。于是所求概率

$$P\{40.5<\bar{\xi}<42\}=\left\{P\frac{40.5-41.5}{\sqrt{0.1173}}<\frac{\bar{\xi}-41.5}{\sqrt{0.1173}}<\frac{42-41.5}{\sqrt{0.1173}}\right\}$$

$$=P\left\{-2.92<\frac{\bar{\xi}-41.5}{\sqrt{0.1173}}<1.46\right\}$$

$$=F_{0,1}(1.46)-F_{0,1}(-2.92)$$

$$=0.9278-0.0017=0.9261。$$

二、样本均值之差

在工商管理工作中，常常要了解两个总体均值之差，在不能直接了解的情况下，通过抽样，以两个样本均值之差去推断两个总体均值之差，这时就需要知道两个样本均值之差（difference of sample averages)的抽样分布。

设从正态总体 $\xi\sim N(\mu_1,\sigma_1^2)$ 中独立地抽取容量为 n_1 的样本，从正态总体 $\eta\sim N(\mu_2,\sigma_2^2)$ 中独立地抽取容量为 n_2 的样本，且这两个样本相互独立，则样本均值之差

$$\overline{\xi}-\overline{\eta}\sim N\left(\mu_1-\mu_2,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right) \tag{1.2.2}$$

这是由于

$$\overline{\xi}\sim N\left(\mu_1,\ \frac{\sigma_1^2}{n_1}\right),\quad \eta\sim N\left(\mu_2,\ \frac{\sigma_2^2}{n_2}\right)$$

所以 $\overline{\xi}-\overline{\eta}$ 也是正态变量，且

$$E(\overline{\xi}-\overline{\eta})=E(\overline{\xi})-E(\overline{\eta})=\mu_1-\mu_2,$$

因为两个样本相互独立，所以 $\overline{\xi}$ 与 $\overline{\eta}$ 也独立，
从而

$$D(\overline{\xi}-\overline{\eta})=D(\overline{\xi})+D(\overline{\eta})=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2},$$

故

$$\overline{\xi}-\overline{\eta}\sim N\left(\mu_1-\mu_2,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)。$$

例 1.2.3 有甲、乙两个煤矿，甲矿平均日产量为 150T，标准差为 20T；乙矿平均日产量为 125T，标准差为 25T。设两个矿的日产量均服从正态分布，从两矿各随机抽取五天的产量计算平均日产量，出现甲矿的平均日产量比乙矿的平均日产量低的可能性有多大？

解 因为两煤矿的日产量 ξ 及 η 都服从正态分布，所以平均日产量 $\overline{\xi}$ 及 $\overline{\eta}$ 也都服从正态分布，$\overline{\xi}-\overline{\eta}$ 也服从正态分布，由于 $\overline{\xi}$ 与 $\overline{\eta}$ 独立，故

$$\overline{\xi}-\overline{\eta}\sim N\left(\mu_1-\mu_2,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)。$$

甲矿的平均日产量低于乙矿的平均日产量，即 $\overline{\xi}-\overline{\eta}<0$，故所求概率

$$P\{\overline{\xi}-\overline{\eta}<0\}=P\left\{\frac{\overline{\xi}-\overline{\eta}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}<\frac{0-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right\}$$

$$= P\left\{\frac{\overline{\xi}-\overline{\eta}-(150-125)}{\sqrt{\frac{20^2}{5}+\frac{25^2}{5}}} < \frac{0-(150-125)}{\sqrt{\frac{20^2}{5}+\frac{25^2}{5}}}\right\}$$

$$= F_{0,1}(-1.75) = 0.0401$$

由此可以看出，出现这种情况的可能性很小，不到 5%。

在许多实际问题中，常常需要从非正态分布的总体进行抽样或从不知其分布类型的总体抽样，为解决样本均值之差的分布类型，就抽取大样本，因为当样本容量很大时，就可以根据中心极限定理来确定样本均值之差的分布。

设两个容量都很大的总体$\overline{\xi}$与$\overline{\eta}$的分布是任意的，总体的均值分别为μ_1与μ_2，总体的方差分别为σ_1^2与σ_2^2。从总体ξ中独立地抽取容量为n_1的样本，从总体η中独立地抽取容量为n_2的样本。当n_1与n_2都很大时，根据中心极限定理样本均值之差$\overline{\xi}-\overline{\eta}$近似服从正态分布

$$N\left(\mu_1-\mu_2,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)。$$

为了求出与统计量$\overline{\xi}-\overline{\eta}$的某些特定值相联系的概率，可以先将$\overline{\xi}-\overline{\eta}$标准化(standardize)，然后查标准正态分布函数$F_{0,1}(x)$数值表。

例 1.2.4 来自某大城市的某住户超过1000户的住宅小区中50户居民调查表，得出每月每户平均生活费为580元，又从另一个住户超过1000户的住宅小区中48户居民调查表，得出每月每户平均生活费为575元。假定此两个住宅小区居民每月每户平均生活费的真值没有差别，且方差都是225元^2，观察到两个样本均值之差$\overline{\xi}-\overline{\eta}$大于或等于5元的概率有多大？

解 两个总体的分布都未知，但样本容量都很大，所以可以引用中心极限定理，又两个样本容量分别相对于各自总体的容量来说又都小于0.05，所以可以不考虑有限总体修正系数，从而统计量

$$\frac{\overline{\xi}-\overline{\eta}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

近似服从标准正态分布 $N(0,1)$，于是所求概率

$$P\{\overline{\xi}-\overline{\eta}\geqslant 5\}=P\left\{\frac{\overline{\xi}-\overline{\eta}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\geqslant\frac{5-0}{\sqrt{\frac{22.5}{50}+\frac{225}{48}}}\right\}$$

$$=P\left\{\frac{\overline{\xi}-\overline{\eta}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\geqslant 1.65\right\}$$

$$=1-F_{0,1}(1.65)=0.0495$$

因此当两个住宅小区的居民每日的平均生活费相等时，观察到两个样本平均生活费之差大于或等于 5 元是不常发生的事。

三、样本比例

从总体中抽取容量为 n 的随机样本，设这个样本中所含的具有某种特征的个体数为 η，称统计量

$$\hat{p}=\frac{\eta}{n} \tag{1.2.3}$$

为样本比例(sample proportion)。

常常希望能够确定样本比例 $\hat{p}$ 大于或等于某个指定值 a 的概率 $P\{\hat{p}\geqslant a\}$，这就需要知道样本比例 $\hat{p}$ 的抽样分布。

$\hat{p}$ 的抽样分布可借助于二项分布 (binomial distribution) 来解决。设总体中具有某种特征的个体比例为 p，则容量为 n 的简单随机样本中含有具有这种特征的个体数 η 服从二项分布 $B(n,p)$。由

$$E(\eta)=np,\qquad D(\eta)=np(1-p)$$