

第四章 参数估计

4.1 数理统计学的基本概念

从本章起,我们转入课程的第二部分——数理统计学.数理统计学与概率论是两个有密切联系的姊妹学科.大体上可以说:概率论是数理统计学的基础,而数理统计学是概率论的重要应用.

数理统计学是一门应用性很强的学科,有其方法、应用和理论基础.在西方,“数理统计学”一词是专指统计方法的数学基础理论那部分而言.在我国则有较广的含义,即包括方法、应用及理论基础都在内,而这在西方是称为“统计学”.在我国,因为还有一门被认为是社会科学的统计学存在,这两个名词的区别使用,有时是必要的.

4.1.1 什么是数理统计学

当我们用试验或观察的方法研究一个问题时,首先要通过适当的观察或试验以取得必要的数据,然后就是对所得数据进行分析,以对所提问题作出尽可能正确的结论.为什么说“尽可能正确”呢?因为数据一般总是带有随机性的误差.需要指出的是,这里指的误差,不单是通常意义下的因测量不准而招致的误差,例如测量一个人的高度,因仪器和操作的原因必然有一定的误差——自然,这种误差也是构成数据的误差的一个可能的来源.这里所说的数据误差,主要指的是由于观察和试验所及,一般只能是所研究的事物的一部分,而究竟是哪一部分则是随机的.例如一个学校有上万名学生,你从中抽出 50 人来研究该校学生的学习情况,抽取的结果(那 50 个人)不同,所得数据就不同,这完全凭机会定.我们说的随机误差主要是指这个.由于数据带有这样的随机性,通过分析这

些数据而作出的结论,也就难保其不出错了.分析方法的要旨,就在于使可能产生的错误愈小愈好,发生错误的机会愈小愈好,这就需要使用概率论的工具,在此我们就可以初步看出概率论和数理统计学的密切关系.

数理统计学就是这样一门学科:它使用概率论和数学的方法,研究怎样收集(通过试验或观察)带有随机误差的数据,并在设定的模型(称为统计模型)之下,对这种数据进行分析(称为统计分析),以对所研究的问题作出推断(称为统计推断).让我们举一个例子来说明这些概念.

例 1.1 某工厂生产大批的电子元件.按第二章例 1.7 的理论,我们认为有理由假定元件的寿命服从指数分布,见第二章(1.20)式.在实际应用中,我们可以提出许多感兴趣的问题.例如:

1. 元件的平均寿命如何?
2. 如果你是使用单位,要求平均寿命能达到某个指定的数 l ,例如 5000 小时.问这批元件可否被接受?

在此,“元件寿命服从指数分布”提供了一个数学模型,即本问题的统计模型(参见例 1.3 中的补充说明).如果你知道了该分布中的参数 λ 之值,则据第三章例 1.4,我们知道平均寿命 $1/\lambda$,于是上面两个问题马上就可以得到回答.但在实用上 λ 往往是未知,于是我们就只好从这一大批元件中随机抽出若干个,例如 n 个,并测出其寿命分别为 X_1, \dots, X_n ,这 n 个元件如何选取?主要是要保证这大批元件中,每一件有同等的被抽出的机会,而这并不是很容易办到的事情,需要想些办法,既能减轻工作量,又能尽可能保证上述同等机会的要求.

有了数据 X_1, \dots, X_n 后,一个自然的想法是:用其算术平均值 $\bar{X} = (X_1 + \dots + X_n)/n$ 去估计未知的平均寿命 $1/\lambda$.当然, \bar{X} 不一定恰好等于 $1/\lambda$.但实际问题中,我们不会也不可能要求所作的估计一丝不差.但误差可能有多大?产生指定大小的误差的机会(概率)有多大?为了使这概率降至指定的限度(例如,0.1),抽出的元件个数 n 至少应达到多少?这些问题的解决方法及有关理论,就

是数理统计学的内容.

本例提出的第一个问题称为参数估计问题,因为 λ 是元件寿命分布中的一个未知参数,而我们的问题是要估计由 λ 决定的一个量,即 $1/\lambda$.也可以把问题提为要求估计参数 λ 本身,这时我们可考虑使用 $1/\bar{X}$ (参见例2.2).参数估计是最重要的统计问题之一.

现在来谈第二个问题.可能认为:至少就本例而言,解决了第一个问题也就解决了第二个问题,因为,既然用 \bar{X} 去估计平均寿命,那就看 \bar{X} 是否不小于指定的数 l .若 $\bar{X} \geq l$,则接受该批产品,不然就不接受.

应当承认,这也是一个可以考虑的解法.但还应注意到,如上文所指出的:因 \bar{X} 估计平均寿命有误差,我们得根据实际需要进行一定的调整.即把接受的准则定为 $\bar{X} \geq l_1$, l_1 是某个选定的数,可以大于、等于或小于 l . l_1 定得大些,表示我们的检验更严格,这在对元件质量要求很高且供货渠道较多时可能是适当的.反之, l_1 定得小些,表示检验更宽,这在对元件质量要求不很高,或急需这些元件而供货渠道很少时,也可能采取.从统计上说,无论你怎么定 l_1 ,理论上你都可能犯两种错误之一:一是元件平均寿命达到需求而被你拒收了,一是元件平均寿命达不到需求而被你接受了.这两种错误各有一定的概率,它们在很大程度上决定了接受准则 $\bar{X} \geq l_1$ 中的 l_1 的选择.

第二个问题与第一个问题不同:它不是要求对分布中的未知参数作出估计,而是要在两个决定(就本问题而言就是接受或拒收该批产品)中选择一个.这类问题称为假设检验问题,也是最重要的统计问题之一.

4.1.2 总体

总体是指与所研究的问题有关的对象(个体)的全体所构成的集合.如在例1.1中,那一大批元件就是问题的总体,而每一个元件就是一个个体,所有这些个体就构成问题的总体.又如:

例 1.2 要研究某大学学生的学习情况,则该校的全体学生构成问题的总体.每一个学生则是该总体中的一个个体.

总体随所研究的范围而定.如在上例中,若你研究全国大学生的学习成绩,则总体就大多了:它包含全国所有在学的大学生.总体如何定,取决于研究目的,也受人力物力时间等因素的限制.

对于大多数实际问题,总体中的个体是一些实在的人或物,而问题中所注意的,并不在于这些人或物本身,而在于所关心的某种指标.例如一个学生有身高体重姓氏笔划籍贯出身……等特征,当我们研究学生学习成绩时,对这些都不关心,而只注意其考分如何.在例 1.1 中,我们只注意元件的寿命如何.这样,也可以把我们感兴趣的那个指标值就作为该个体(例如,大学生 A 得 90 分,即以 90 这个数代替 A),而总体就由一些数所组成.

单是这样还不行.这里有两个问题:一是总体中这样一大堆杂乱无章的数没有赋予什么数学或概率的性质,因而无法使用有力的概率论工具去研究它;二是各种总体变得没有区别.例如,大学生的学习成绩也是一堆数,一大批元件的寿命也是一堆数,大家都一样了.解决这些问题的途径,就涉及总体这个概念的核心——总体的概率分布.例如,在例 1.1 中元件寿命分布为指数分布,例 1.2 学生的学习成绩可以假定为服从正态分布.总体分布不同,分析的方法也就不同,赋有一定概率分布的总体就称为统计总体.

因此,经过以上几步的分析,我们就得出在数理统计学中“总体”这个基本概念的要旨——总体就是一个概率分布.当总体分布为指数分布时,称为指数分布总体;当总体分布为正态分布时,称为正态分布总体或简称正态总体,等等.两个总体,即使其所含个体的性质根本不同,只要有同一的概率分布,则在数理统计学上就视为是同类总体.例如人的寿命也可以服从指数分布,它与元件寿命的分布一样,处理二者的统计问题的方法也一样,即可视为同一类总体.

对以上所说的要作一点说明:如例 1.1 所显示的,虽然我们假定了元件寿命服从指数分布,但并没有指定其中参数 λ 之值.既

然 λ 未知,原则上 λ 可取 0 到 ∞ 内任何值,故更正确地应当说:总体分布是一个概率分布族(在此为指数分布族)的一员.这分布族包含一个参数 λ ,称为单参数分布族.例 1.2 的总体分布——正态分布 $N(\mu, \sigma^2)$,包含两个参数 μ 和 σ^2 (μ 可取任何实数值而 σ^2 只能取大于 0 的值),是一个两参数分布族.另外,在有些情况下,我们只是假定总体有一定的概率分布而并不明确知道其数学形式.如在例 1.1 中,也可以只承认寿命有一定的概率分布函数 $F(x)$, $F(0)=0$ (因寿命总大于 0),其他别无所知.这时,总体分布不能通过若干个未知参数表达出来,这种情况称为非参数总体.对非参数总体,虽不知其数学形式,但统计问题照样可以提出来.例如估计平均寿命的问题,不假定元件寿命分布为指数分布也有意义,且使用 \bar{X} 去估计平均寿命看来仍是一个合理的方法.自然,由于分布的形式未知,进一步的讨论困难就更大,这些在以后会逐步指明.

上面所讲的总体概念,在很大程度上要归功于数理统计学最主要的奠基者,伟大的英国统计学家 R. A. 费歇尔.他引进了“无限总体”这个概念——现实问题中,当所考察的个体是由一些看得见、摸得着的对象所构成时(如例 1.1, 1.2),总体总是有限的.有限总体相应的分布只能是离散的,其具体形式将与个体总数有关且缺乏一个简洁的数学形式,这会使有力的概率方法无法使用.引进无限总体的概念,在概率论上相当于用一个连续分布去逼近离散分布.当总体所含个体极多时,这种逼近所带来的误差,从应用的观点看已可以忽略不计.更好的是,事实证明:几种常见且在概率论上较易处理的分布,如指数分布和正态分布等,尤其是正态分布,对许多实用问题的总体分布给出了足够好的近似,而围绕着这些分布建立了深入而有效的统计方法.

最后,关于总体这个概念还需要说明一个问题.从一个例子入手,设有一个物体,其真实的重量 a 未知,要通过多次量测的结果去估计它.请问在这个问题中总体是什么?若不假思索,可能回答说:因为与所研究的问题有关的对象,就只这个物体,故这个物体,或者其重量 a ,就构成总体,这个回答不对.其所以不对,一则因为

a 未知. 即使 a 已知(这时自然不存在估计它的问题, 但量测其重量仍有意义, 例如, 可能是为了考察天平的准确程度如何), 这个回答仍不对. 因为你既然通过量测, 那么, 你所研究的问题, 实质上是“通过量测结果去估计 a 之值其精度如何”. 这样, 每一个可能的量测结果都是一个个体, 而总体是由“一切可能的量测结果”组成. 这只是一个想像中存在的集合, 因为不可能去进行无限次量测, 把所有可能的量测结果一一列出来. 这与我们前面几个例子中那种看得见摸得着的总体不同: 这里的总体只是在想像中存在, 它的个体是通过试验“制造”出来的——每秤一次, 就制造出一个量测值. 这种情况在实际应用中非常之多. 给这种总体规定分布也一样. 拿本例来说, 只须说一句“量测结果服从某某分布(如正态分布)”就行. 如果不绕这么一个圈子, 而直接说: 量测结果是随机的, 它服从某某分布, 可能读者会感到更易接受. 上述分析是为了突出统计总体这个概念的这种抽象形式, 以体现这个概念的普遍性.

在某些统计学著作中, 也常把总体称为“母体”.

4.1.3 样本

样本是按一定的规定从总体中抽出的一部分个体. 所谓“按一定的规定”, 就是指总体中的每一个个体有同等的被抽出的机会, 以及在这个基础上设立的某种附加条件.

由于我们的兴趣不在于个体本身而在于其某一特征指标值, 所得样本表现为若干个数据 X_1, \dots, X_n . n 称为“样本大小”或“样本容量”, “样本量”. 样本 X_1, \dots, X_n 中的每一个 X_i 也称为样本. 有时, 为区别这种情况, 把 X_1, \dots, X_n 的全体称为一“组”样本, 而 X_i 称为其中的第 i 个样本.

在一个具体问题中, 样本 X_1, \dots, X_n 是一些具体的数据. 而在理论的研究上, 则要把它看成为一些随机变量. 因为抽到哪一些个体是随机的, 因而其指标值, 即 X_1, \dots, X_n , 也是随机的.

设想样本是一个一个地抽出来. 第一次抽时, 是从整个总体中抽一个. 因而 X_1 的分布也就与总体分布相同. 如果这一个不放回

去,到第二次抽时,总体中已少了一个个体,其分布有了变化,因此 X_2 的分布会与 X_1 的分布略有差别.但是,如果总体中所包含的个体极多,或如理论上设想的,总体中包含无限多个个体,则抽掉一个或几个,对总体的分布影响极少或毫无影响.这时, X_1, \dots, X_n 独立且有相同的分布,其公共分布即总体分布.这是在应用上最常见的情形,也是理论上研究得最深入的情形,本节主要考虑这种情况.在数理统计学上,称这种情况为: X_1, \dots, X_n 是从某总体中抽出的独立随机样本,或简称为从某总体中抽出的样本.

当总体中所含个体数不太大时,情况就不同.考察以下的例子:

例 1.3 设一批产品包含 N 个,内有废品 M 个, M 未知.因而废品率也未知.现从其中抽出 n 个逐一检查它们是否为废品,据此去估计 p .

如果把合格品记为0而废品记为1,则总体分布为离散分布 $P(X=1)=p, P(X=0)=1-p$.设想样本是一个一个抽出的,结果记为 X_1, \dots, X_n .如果抽样是有放回的,即每抽出一个作检查以后再放回去,下次仍有同等机会被抽,则 X_1, \dots, X_n 为独立同分布,每一个的分布就是上述总体分布.若用 $\bar{X}=(X_1+\dots+X_n)/n$ (即样本中的废品率)去估计 p ,则因 $X_1+\dots+X_n$ 服从二项分布 $B(n, p)$ (见第二章例1.1).这个估计的统计性质就由此决定了.

另一种抽样方式,即常见的作法,是一次抽出 n 个或一个一个抽但已抽出的不再放回.这时,用 \bar{X} 估计 p 仍是一个合理的选择,但因 $X_1+\dots+X_n$ 已不是二项分布而是超几何分布(见第二章例1.4).这个估计的统计性质就与上面所讲的有所不同.当 N 不很大时,这个差别不可忽视.

由此例可见,在有限总体的情况,单由总体分布已不足以完全决定样本的分布如何,要看抽样的方式.这样,抽样的方式也要作为一个要素加入到统计模型的内容中来.在无限总体的情况,或者是有限总体而抽样有放回的情况,按第二章定义3.1,总体分布完全决定了样本的分布,故就可以把总体分布等同于统计模型.

4.1.4 统计量

完全由样本所决定的量,叫做统计量.这里要注意的是“完全”这两个字.它表明:统计量只依赖于样本,而不能依赖于任何其他未知的量.特别是,它不能依赖于总体分布中所包含的未知参数.

例如,设 X_1, \dots, X_n 是从正态总体 $N(\mu, \sigma^2)$ 中抽出的样本,则 $\bar{X} = (X_1 + \dots + X_n)/n$ 是统计量,因为它完全由样本 X_1, \dots, X_n 决定. $\bar{X} - \mu$ 不是统计量,因为 μ 未知, $\bar{X} - \mu$ 并不完全由样本所决定.

统计量可以看作是对样本的一种“加工”,它把样本中所含的(某一方面的)信息集中起来.例如,上述 \bar{X} 可用于估计未知的 μ .可以这样看:原始数据 X_1, \dots, X_n 中的每一个,都包含有 μ 的若干信息,但这些是杂乱无章的,一经集中到 \bar{X} ,就有了更明确的概念.所以,有用的统计量都是“有的放矢”的,针对某种需要而构造的.如在上例中,若想了解有关总体方差 σ^2 的情况,则统计量 \bar{X} 没有什么用.从方差是反映散布度这方面去看,下面的统计量

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1) \quad (1.1)$$

是有用的.因为 S^2 是样本 X_1, \dots, X_n 的散布程度的一个合理的刻画,它应当与 σ^2 有密切的关系, S^2 这个重要的统计量叫做“样本方差”.

有一类重要的统计量叫做样本矩,分为样本原点矩和样本中心矩.设 X_1, \dots, X_n 为样本, k 为正整数.则

$$a_k = (X_1^k + \dots + X_n^k) / n \quad (1.2)$$

称为 k 阶样本原点矩. $a_1 = \bar{X}$ 是最重要的样本原点矩,它常称为“样本均值”.而

$$m_k = \sum_{i=1}^n (X_i - \bar{X})^k / n \quad (1.3)$$

称为 k 阶样本中心矩.

在第三章定义 2.2 中,我们定义过随机变量 X 的 k 阶原点矩

a_k 和 k 阶中心矩 μ_k . 此处定义的 a_k, m_k 是它们的样本对应物. 有时也把 a_k 和 μ_k 称为理论矩, 而 a_k, m_k 称为经验矩. 这名词可以用如下的方式去解释: 设总体分布 F 有(理论)矩 a_k, μ_k . 由于不知道 F , 也就不知道 a_k, μ_k . 现在有从该总体中抽出的样本 X_1, \dots, X_n , 我们就构造一个分布 F_n 去模拟 F . 由于手头这 n 个样本 X_1, \dots, X_n 的地位是平等的, 一个合理的选择是把 F_n 取成一个离散分布, 它在每个值 X_i 处各有概率 $1/n, i = 1, \dots, n$. 形式地, 分布函数 F_n 定义为

$$F_n(x) = \{X_1, \dots, X_n \text{ 中 } \leq x \text{ 的个数}\} / n \quad (1.4)$$

它称为样本 X_1, \dots, X_n 的经验分布函数. 如果按第三章定义 2.2 计算分布 F_n 的 k 阶原点矩和中心矩, 则分别得到 a_k 和 m_k . 所以, 样本矩无非就是经验分布的矩.

特别值得注意的二阶中心矩 m_2 . 它与样本方差 S^2 只相差一个常数因子: $m^2 = \frac{n-1}{n} S^2$.

最有用的样本矩是一、二阶的, 三、四阶的也有些应用. 四阶以上的则很少使用.

有用的统计量很多, 它们都是在解决种种统计推断问题时产生的, 以后将结合这些问题来介绍.