第三章　统计推断及其常见错误

前面简单介绍了统计的过程,知道对于一个统计试验,通过对收集到的数据进行加工整理,然后进行统计分析,就可以形成对总体数量规律和基本认识。但在实际工作中,所能收集到的资料绝大部分是样本资料,而不是总体的全部资料,即使能够收集到总体的全部资料,也会因为统计成本过高而拒绝这样做。这就要求我们在数据资料有限的情况下,通过对一定的样本信息的处理、分析,来做出对总体规律的基本推断。

第一节　抽样推断法

一、抽样推断的理论基础

从原理上讲,要证明统计规律适用于整个总体,方法之一就是对总体的每一个单位都进行调查,进而获取总体的全部资料,但这种方法在实践中常常是不现实的。原因是:

（1）总体的某些部分在调查时可能还不存在。如，要了解某种药物作用于某种疾病上的效果，可能会缺乏所需的病例。

（2）试验本身可能是破坏性的。如要想了解一批汽车的抗撞击的能力，必须对其进行撞击实验，但试验的结果可能会导致全部汽车的报废。

（3）研究者可能受到时间、人力、物力、财力等条件上的限制。

由于以上原因，通常有必要借助于总体中的部分资料，即样本，来得到适用于总体的信息。实际应用中，样本可分为代表性样本、有偏样本、随机样本、分层样本和整群样本等类型，根据实际问题的需要，可以选择适当类型的样本。而进行抽样推断一般使用简单随机样本。

设$(X_1,X_2,\cdots,X_n)$是来自总体X的样本，如果它满足如下两个条件，则称为简单随机样本：①$X_1,X_2,\cdots,X_n$与X同分布；②$X_1,X_2,\cdots,X_n$彼此独立。由于简单随机样本的分量具有独立同分布的特征，这使得利用样本或统计量对总体统计特性进行统计分析变得简单化。因为此时样本分布与总体分布有着简单的关系：$F_n(x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} F(x_i)$。其中$F_n$和$F$分别是样本和总体的分布函数。

由于统计推断是由局部认识整体的一个过程，一个自然的想法，就是要求样本要有代表性，即要求每个个体被选中的机会均等，并且抽取一个个体后总体成分不变。首先要求抽样具有“随机性”，第一次抽取的样品X_1的可能取值与总体的可能取值是完全一样的，且取各个值的概率相同。因此X_1是一个随机变量，并且是与X同分布的随机变量。其次应具有“独立性”，第一次的抽样不改变总体成分，第二次抽取的样品X_2的可能取值也与总体的可能取值完全一样，并且取值的概率也相同，因此X_2也是与X同分布的一个随机变量，且与X_1是相互独立的。同理，$X_3,X_4,\cdots,X_n$都是与X同分布的随机变量，并且$X_1,X_2,\cdots,X_n$是一组相互独立的随

机变量，故要求 $X_3, X_4, \cdots, X_n$ 是简单随机样本。此外，要求样本是简单随机样本，可以更好地应用概率论条件下的一系列结论，而这些结论恰是统计学所必要的理论基础[6]。现在的问题是，在实践中如何获得简单随机样本?从理论上讲，对总体进行随机地、独立地重复观测，便可以获得简单随机样本。这里抽样的随机性是指对总体的每一个个体有相同的机会被抽取，抽样的独立性是指每次抽样结果发生的可能程度不受其他抽样结果的影响。对于从一个有限总体抽取样本时，随机有放回地抽取便可以保证样本分量的独立性，而当总体所含个体数目大而所抽取样本的容量相对比较小时，尽管抽取是无放回的，所得样本的分量也可近似看成是相互独立的。由概率论的知识可知，对于固定的样本容量，当总体所含个体数目趋于无穷时，有放回抽取和无放回抽取所得到样本的分布将趋于一致。当抽取样本容量 n 相对于总体数 N 很小时(一般 $n/N \leqslant 10$)，则连续抽取 n 个个体，就近似地看作一个简单随机样本，这是因为抽取的个数很小时，可以认为对总体不影响或影响很小的缘故。如果采取有放回抽样，则不必要求 n 相对很小。

显然，按上述方法获取的样本对总体具有代表性。但是，尽管简单样本在推断统计中具有很多优点，但因为抽样推断是在样本数据的基础上进行的，所以推断结果产生误差是难免的。我们的任务就是如何利用数学工具，去测定误差的范围，并对误差进行控制。由于样本只能提供一些零散的、杂乱的数量信息，必须对样本数据进行整理、汇总，计算样本值某些统计指标，并由这些统计指标对总体进行预测或判断。实际中经常用到的统计量有以下几种：

(1) 样本均值：$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$；

(2) 样本方差：$S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$；

(3) 修正样本方差：$S_n^{*2} = \dfrac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$；

(4) 样本 k 阶原点矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1,2,\cdots$；

(5) 样本 k 阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k, \quad k = 1,2,\cdots$。

统计量的概率分布称为抽样分布。通常情况下，概率分布可以用均值和方差这两个重要的数学特征值概括地描述。由于实际上不可能也不必要从同一总体中抽取所有可能样本，而总是抽取一个样本据以推断的，因此抽样分布是统计推断的基础。我们总是试图依据抽样分布的理论，对总体参数进行估计。下面是几个常用统计量的抽样分布。

（一）样本均值的抽样分布

1. 单个正态总体下的样本均值的分布

设总体 X 服从正态分布 $N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 是来自 X 的一个样本，则样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 服从均值为 μ，方差为 $\frac{\sigma^2}{n}$ 的正态分布，即 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu,\frac{\sigma^2}{n})$。这说明均值取值平均大小与总体 X 取值平均大小相等。但当 n 增大时，均值取值关于数学期望的集中程度远比总体要高，当 $n\to\infty$ 时，$\bar{X}$ 取值几乎集中在数学期望 μ 这点处。

2. 两个正态总体下的样本均值的分布

定理：设有两个正态总体 $X \sim N(\mu_1,{\sigma_1}^2)$，$Y \sim N(\mu_2,{\sigma_2}^2)$ 相互独立，$(X_1,X_2,\cdots,X_n)$ 和 $(Y_1,Y_2,\cdots,Y_n)$ 为分别来自 X 和 Y 的简单随机样本，$\bar{X}$，$\bar{Y}$ 分别为其样本均值，则

$$\bar{X} \pm \bar{Y} \sim N(\mu_1 \pm \mu_2, \frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n})$$

3. 非正态总体下的样本均值的近似分布

设 X 为任意总体，其数学期望为 $E(X) = \mu$，方差为 $D(X) =$

σ^2,又 $X_1,X_2,\cdots,X_n$ 是来自 X 的一个样本,当 n 较大时,近似地有 $\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$。更一般地,有

定理:设总体 X(不管服从什么分布,只要均值和方差存在)的均值为 μ,方差为 σ^2,$X_1,X_2,\cdots,X_n$ 为 X 的一个样本,则 $E(\overline{X}) = \mu$, $D(\overline{X}) = \frac{\sigma^2}{n}$,$E(s^2) = \sigma^2$。

(二) χ^2 分布

设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立,同服从标准正态分布 $N(0,1)$,则随机变量 $\chi^2 = X_1{}^2 + X_2{}^2 + \cdots + X_n{}^2$ 的概率密度函数为

$$f(x) = \begin{cases} \dfrac{1}{2^n\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

称随机变量 χ^2 服从自由度为 n 的 χ^2 分布,记为 $\chi^2 \sim \chi^2(n)$。这里的自由度是指上式右端包含的独立变量的个数。

因 $f(x)$ 的表达不简洁,直接用它计算概率较困难,为方便计,根据 $P\{\chi^2(n) > \chi_\alpha{}^2(n)\} = \alpha(0 < \alpha < 1)$ 作成 χ^2 的分布表供查阅。一般,对于给定的 $0 < \alpha < 1$,称满足条件 $\int_{\chi_\alpha^2(n)}^{+\infty} f(x)\mathrm{d}x = \alpha$ 的横坐标 $\chi_\alpha^2(n)$ 为 χ^2 上的上侧 α 分位点(分位数)。χ^2 分布对参数 n 具有可加性,即若随机变量 X 和 Y 分别服从自由度为 n_1 和 n_2 的 χ^2 分布,且相互独立,则 $X + Y$ 服从自由度为 $n_1 + n_2$ 的 χ^2 分布。

(三) t 分布

设随机变量 X 和 Y 相互独立,X 服从标准正态分布 $N(0,1)$,Y 服从自由度为 n 的 χ^2 分布,则随机变量 $t = \frac{X}{\sqrt{Y/n}}$ 的概率密度函数为

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ -\infty<x<+\infty$$

称随机变量 t 服从自由度为 n 的 t 分布,记为 $t\sim t(n)$。可以证明,当 $n\to+\infty$ 时,t 分布的极限是标准正态分布。事实上,当 $n>30$ 时,t 分布和标准正态分布的概率密度曲线几乎相同。当 n 较小时,t 分布和标准正态分布相差很大。

(四)F 分布

设随机变量 X 和 Y 相互独立,$X\sim\chi^2(n_1)$,$Y\sim\chi^2(n_2)$,则随机变量 $F=\left(\frac{X}{n_1}\right)\Big/\left(\frac{Y}{n_2}\right)$的概率密度函数为

$$f(x)=\begin{cases}\dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}\dfrac{x^{\frac{n_1}{2}-1}}{(n_1x+n_2)^{\frac{n_1+n_2}{2}}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

随机变量 F 服从自由度为(n_1,n_2)的 F 分布,记为 $F\sim F(n_1,n_2)$。其中 n_1 是分子的自由度,称为第一自由度,n_2 是分母的自由度,称为第二自由度。给定 $0<\alpha<1$, 使

$$P(F>F_{1-\alpha}(n_1,n_2))=\int_{F_{1-\alpha}(n_1,n_2)}^{+\infty}f(x)\mathrm{d}x=\alpha$$

称 $F_{1-\alpha}(n_1,n_2)$ 为 $F(n_1,n_2)$ 分布的上侧 $100\alpha\%$ 分位点。F 分布函数表,一般取 α 为 0.25, 0.10, 0.05, 0.025, 0.01 等。上侧分位数都是大于 1 的数,当需要小于 1 的分位数时,可以由 F 分布的性质得到。

二、抽样分布的应用技巧

定理: 设 $X_1,X_2,\cdots,X_n$ 与 $Y_1,Y_2,\cdots,Y_n$ 分别是来自正态总体 $N(\mu_1,{\sigma_1}^2)$ 和 $N(\mu_2,{\sigma_2}^2)$ 的样本,且这两个样本相互独立。设 $\overline{X}=$

$\frac{1}{n_1}\sum_{i=1}^{n_1}X_i, \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2}Y_i$, $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$, $S_2^{\,2} = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$, $S_w^{\,2} = \frac{(n_1-1)S_1^{\,2}+(n_2-1)S_2^{\,2}}{n_1+n_2-2}$, 则有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

统计中之所以引入S_w,是因为在定理的条件下,有$E(S_1^{\,2}) = \sigma^2$和$E(S_2^{\,2}) = \sigma^2$,即$S_1^{\,2}$和$S_2^{\,2}$都可作为σ^2的估计,但它们都只利用了部分信息,将它们合起来估计σ^2会更好。因为$\frac{(n_1-1)S_1^{\,2}+(n_2-1)S_2^{\,2}}{\sigma^2} \sim \chi^2(m+n-2)$,因此,若令$S_w^{\,2} = \frac{(n_1-1)S_1^2+(n_2-1)S_2^{\,2}}{n_1+n_2-2}$,仍有$E(S_w^{\,2}) = \sigma^2$。

(一)正态总体下常用统计量的分布

正态总体下常用统计量的分布是区间估计和假设检验的基础,在应用中要注意以下差别:

(1)公式中$\sqrt{n}(\bar{X}-\mu)/\sigma \sim N(0,1)$,如果$\sigma$未知,可用$S$代替,此时,$N(0,1)$要改为$t(n-1)$,即$\sqrt{n}(\bar{X}-\mu)/S \sim t(n-1)$;

(2)公式$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$中,如果$\mu$未知,则用$\bar{X}$代替,此时,$\chi^2(n)$应换成$\chi^2(n-1)$,即$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$;

(3)公式$\dfrac{\frac{1}{m\sigma_1^{\,2}}\sum_{i=1}^{m}(X_i-\mu_1)^2}{\frac{1}{n\sigma_2^{\,2}}\sum_{i=1}^{n}(Y_i-\mu_2)^2} \sim F(m,n)$中,在$\mu_1$,$\mu_2$未知的情况下,可用$\bar{X}$, $\bar{Y}$代替, 此时, m, n应分别换成$m-1$和$n-1$;

(4) 在公式$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{{\sigma_1}^2}{m}+\dfrac{{\sigma_2}^2}{n}}} \sim N(0,1)$中，如果${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$未知，把$\sigma$提出根号外并用$S_w$代替，则有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$$

(二) χ^2分布、t分布、F分布及正态分布间的常见关系

(1) 若$X_1, X_2, \cdots, X_n$是独立同分布于正态分布$N(0,1)$，则$\sum\limits_{i=1}^{n} {X_i}^2 \sim \chi^2(n)$；

(2) 如果$X \sim N(0,1)$，$Y \sim \chi^2(n)$且X, Y相互独立，则$\dfrac{X}{\sqrt{Y/n}} \sim t(n)$；

(3) 如果$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$且$X, Y$相互独立$\dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$，而$X+Y \sim \chi^2(n_1+n_2)$；

(4) 如果$T \sim t(n)$，则$T^2 \sim F(1,n)$；

(5) 如果$F \sim F(n_1, n_2)$，则$\dfrac{1}{F} \sim F(n_1, n_2)$；

(6) 如果${\chi_n}^2 \sim \chi^2(n), n=1,2,\cdots$，则当$n \to \infty$时，$\dfrac{{\chi_n}^2-n}{\sqrt{2n}}$渐近服从正态分布$N(0,1)$；

(7) 如果$T_n \sim t(n), n=1,2,\cdots$，则当$n \to \infty$时，$T_n$渐近服从正态分布$N(0,1)$；

(8) 如果$Z=\ln X, X \sim F(n_1, n_2)$，则当$n \to \infty$时，$Z$渐近服从正态分布$N\left(\dfrac{1}{2}\left(\dfrac{1}{n}-\dfrac{1}{m}\right), \dfrac{1}{2}\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\right)$；

(9) 如果 $X \sim F(n_1, n_2)$，则 $n \to \infty$ 时，X 渐近服从于 $\frac{1}{n_1}\chi_{n_1}^2$ 的分布，其中 ${\chi_{n_1}}^2 \sim \chi^2(n_1)$。

第二节　参数估计

一、点　估　计

如果对随机样本的数据信息进行整理，得到一个估计量，以这个估计量去估计总体的一个参数，那么，这种单一估计的方法就称为点估计。具体来说是，总体 X 的分布函数的形式是已知的，但它的一个或多个参数未知，可以借助于总体 X 的一个样本来估计总体未知参数的值。点估计的理论依据是大数定理。由大数定理知，无论总体的分布如何（只要有有限方差），样本均数的分布随着样本容量的增加，将越来越集中在总体均数附近。因此样本越大，越能肯定样本均数是总体均数的一个较好的估计。

常用的点估计方法有三种：顺序统计量估计法、矩估计法和极大似然估计法。这三种估计的数学定义相对复杂，限于篇幅，在此不便陈述，大家可以参阅相关的数理统计教材。顺序统计量估计法因不需要太多计算，所以应用最为方便，但这种方法要求总体服从正态分布，或者总体的分布是对称的，而且计算的精度不高。矩估计法直观意义最明显，其做法是：以样本矩作为总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量，从而得到总体未知参数的估计。矩估计法对任何总体都可以，方法简单，但要求总体的相应矩存在，若不存在，就不能用矩估计法。极大似然估计法的基本思想是，如果已观察到样本 $(X_1, X_2, \cdots, X_n)$ 的样本值 $(x_1, x_2, \cdots, x_n)$，而取到这一样本值的概率为 P（离散型的情况），或 $(X_1, X_2, \cdots, X_n)$ 落在这一样本值 $(x_1, x_2, \cdots, x_n)$ 的邻

域内的概率为P(连续型的情况),而P与示知参数θ有关,我们就取θ的估计值使概率P取到最大。极大似然估计法也是对任何总体都可以用,从它得到的估计量具有一致性和有效性,即使不具有无偏性,也常常能够修改成无偏估计量,可以证明,在一定条件下,未知参数的极大似然估计量与其真值之差可以任意小。所以,从某种意义上说没有比极大似然更好的估计。在统计问题中也是往往先使用极大似然估计法。但是需要强调的一点就是,并不是所有待估计的参数都能求到似然估计量,并且在求极大似然估计量时,往往要解一个似然方程(或方程组),有时这种方程非常难解甚至根本写不出有限形式的解,此种情况下,必须考虑使用矩估计法等其他方法。另外,需要说明的一点,用矩估计法和用极大似然估计法所得到的估计有时并不相同。例如,设总体$X \sim N(\mu,\sigma^2)$,μ,σ为未知参数,$(X_1,X_2,\cdots,X_n)$为来自总体X的样本,可以计算验证,μ和σ^2的矩估计和极大似然估计相同,都分别为$\overline{X}$和$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$。又如,总体$X \sim U(0,\theta)$,其中$\theta>0$为未知参数,$(X_1,X_2,\cdots,X_n)$为来自总体$X$的样本,由计算知$\theta$的矩估计为$2\overline{X}$,$\theta$的极大似然估计为$\max\limits_{1\leqslant i\leqslant n}\{X_i\}$。

在点估计中,矩估计和极大似然估计是两种最常用的形式。对于同一个参数,用不同的方法求得的估计量可能是不同的,我们常用无偏性、有效性和一致性(又称为相合性)作为评选估计量的三个标准。

运用样本数据对总体未知参数θ作估计时,总希望这个估计尽可能准确和有效,但在评价估计量时,由于估计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$是样本的函数,它是一个随机变量,对于不同的样本观察值会得到不同的参数估计值,因此不能仅由一次试验的结果来衡量。我们首先希望多次估计值的理论平均等于真值θ,即$E(\hat{\theta})=\theta$,这一

性质称为无偏性。其次，对于同一个参数，可能有许多无偏估计量，究竟哪一个估计量是最好的呢?我们自然认为应以对真值的平均偏差较小者为好，即若$\hat{\theta}_1$和$\hat{\theta}_2$都是θ的无偏估计量，如果$D(\hat{\theta}_1) < D(\hat{\theta}_2)$则称$\hat{\theta}_1$比$\hat{\theta}_2$有效。最后，估计量$\hat{\theta}$的无偏性和有效性都是在样本容量$n$愈大时估计量$\hat{\theta}$愈接近被估计参数，从而引出了估计量的一致性，即对任意的$\varepsilon > 0$，有$\lim\limits_{n\to\infty}P\{|\hat{\theta}-\theta| > \varepsilon\} = 0$，称$\hat{\theta}$是$\theta$的一致估计量。由于一致性是在极限意义下引进的，因而，只有样本容量相当大时，才能显示出其优越性，这在实际中往往难以做到，而且在有些情况下，证明估计量的一致性并非容易，因此，在实际工作中常常使用无偏性和有效性这两个指标。

二、区 间 估 计

在点估计中，是以样本资料去求得一个样本统计量的值，然后以该值作为总体参数的估计值，因此，点估计是因样本而异的。实际情况是，很少有一个样本能充分地提供对总体特征的完全正确的估计，即便是最佳估计量，也有可能与真值有一定的误差。能够进一步知道近似值的精确程度(即所求真值所在的范围)是人们在测量或计算时最需要和最希望的。因此，对于未知参数θ，除了求出它的点估计外，还希望估计出一个范围，并希望知道这个范围包含参数真实值的可信程度。这样的范围通常以区间的形式给出，同时还给出此区间包含参数θ真值的概率保证。这种形式的估计称为区间估计，这样的区间即所谓的置信区间[6]。其数学定义是：设总体X的分布函数$F(x;\theta)$含有一个未知参数θ，$\theta \in \Theta$(是θ可能取值的范围)，对于给定值$\alpha(0 < \alpha < 1)$，若由来自X的样本$X_1, X_2, \cdots, X_n$确定的两个统计量$\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n)$和$\bar{\theta} = \bar{\theta}(X_1, X_2, \cdots, X_n)$ $(\theta < \bar{\theta})$，对于任意的$\theta \in \Theta$满足

$$P\{\underline{\theta}(X_1,X_2,\cdots,X_n) < \theta < \bar{\theta}(X_1,X_2,\cdots,X_n)\} \geqslant 1-\alpha$$

则称随机区间$(\underline{\theta},\bar{\theta})$是$\theta$的置信水平为$1-\alpha$的置信区间,$\underline{\theta}$和$\bar{\theta}$分别称为置信水平为$1-\alpha$的双侧置信区间的置信下限和置信上限。上式的含义是:若反复抽样多次(各次得到的样本容量相等,都是n),每个样本确定一个区间$(\underline{\theta},\bar{\theta})$,每个这样的区间要么包含$\theta$的真值,要么不包含$\theta$的真值,由大数定理,在上述的区间中,包含$\theta$真值的约占$100(1-\alpha)\%$,不包含$\theta$真值的约占$100\alpha\%$。$P\{\underline{\theta} < \theta < \bar{\theta}\} = 1-\alpha$与$P\{a < \theta < b\} = 1-\alpha$的主要区别是,后者中区间$(a,b)$是一个确定的数值区间,$\theta$是一个随机变量,它表示$\theta$的取值落在区间$(a,b)$内的概率是$1-\alpha$;而$P\{\underline{\theta} < \theta < \bar{\theta}\} = 1-\alpha$中的区间$(\underline{\theta},\bar{\theta})$是随机区间,$\theta$是客观存在的一个未知常数,它的含义是随机区间$(\underline{\theta},\bar{\theta})$包含常数$\theta$的概率是$1-\alpha$。例如,若$\alpha=0.05$,反复抽样1000次,每一组样本观察值代入到$\underline{\theta}(X_1,X_2,\cdots,X_n)$和$\bar{\theta}(X_1,X_2,\cdots,X_n)$中就得到一个随机区间,1000个样本就得到1000个随机区间,那么在其中置信水平为0.95的置信区间里,平均有950个包含真值θ,只有50个不包含真值θ,就是说可能有5%的置信区间被误认为包含真值θ而犯了错误。换句话说,就是当给出参数θ的置信水平为0.95的置信区间$(\underline{\theta},\bar{\theta})$时,可以有95%的把握认为区间$(\underline{\theta},\bar{\theta})$包含真值$\theta$,犯错误的概率为0.05。因此,置信水平反映的是区间$(\underline{\theta},\bar{\theta})$包含真值$\theta$的可靠程度,显然这种可能性越大越好。另一方面,还希望估计的精确度高。例如估计某产品的次品要在(0,0.99)内,这个估计虽然有很高的可靠性,但其精确度太差,以至于这种估计毫无意义。区间估计就是寻找这种方法,以构造出具有较高的可靠性和精确度的区间,但是,当样本容量固定不变时,置信区间的可靠性越高(即$1-\alpha$越大),区间估计的精确度就越差(区间的长度就越长),所以在n固定时,置信区间的可靠性和精确度是相互制约的,不可能同时将两者提高到

任意的高度，从置信区间的结构可以看出，增大样本容量，可以缩短置信区间的长度。求置信区间，实际上就是在保证可靠性达到指定水平 $1-\alpha$，犯错误的概率控制在 α 之内的前提下，尽可能地提高精确度[6]。

这里需要强调的一点是，对于同一置信度，置信区间并不惟一。例如，设总体 $X \sim N(\mu, 0.4^2)$，样本容量 $n=20$，样本均值的观察值 $\bar{X}=32.3$，求 μ 的置信度为 $1-\alpha=0.95$ 的置信区间。在此，选 $U=\dfrac{\bar{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}} \sim N(0,1)$，由 $P\{\mu_1<\mu<\mu_2\}=1-\alpha$，即

$P\left\{\bar{X}-\mu_2\dfrac{\sigma}{\sqrt{n}}<\mu<\bar{X}-\mu_1\dfrac{\sigma}{\sqrt{n}}\right\}=1-\alpha$，及 $n=20$，$\sigma=0.4$，解得

μ 的置信度为 $1-\alpha$ 的置信区间为 $\left(\bar{X}-\mu_2\dfrac{\sigma}{\sqrt{n}}, \bar{X}-\mu_1\dfrac{\sigma}{\sqrt{n}}\right)$。显然 μ_1 和 μ_2 在可能的范围内取不同的值，就能得到不同的置信区间。如取 $\mu_1=u_{\frac{\alpha}{2}}=u_{0.025}$，$\mu_2=u_{1-\frac{\alpha}{2}}=u_{0.975}$，查标准正态分布表知 $\mu_1=-1.96$，$\mu_2=1.96$，于是得 μ 的一个置信区间为(32.12，32.48)。上面是将显著水平 $\alpha=0.05$ 分成两个相等的部分，如果将它分成不相等的两部分，如 $\alpha_1=0.01$，$\alpha_2=0.04$，$\mu_1=u_{0.01}$，$\mu_2=u_{1-\alpha_2}=u_{0.96}$，可算得 μ 的另一个置信区间为(32.14，32.51)。容易看出，不同的置信区间长度一般不相等。可以证明，假如构造置信区间的函数具有对称分布（如正态分布，t 分布等）时，以样本均值的观察值为中心的对称区间，其长度最短，误差范围小，估计精度高。下面介绍几种常用的区间估计。

（一）σ^2 已知时总体均值的区间估计

若总体服从正态分布 $N(\mu,\sigma^2)$，则 $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$。令 $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$，则可以将非标准正态分布转化为标准正态分布，即 $Z \sim$

$N(0,1)$。对于给定的置信度 $1-\alpha$，可以查正态分布表，得出相应的临界值 $Z_{\frac{\alpha}{2}}$ 使得 $P\{-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\} = 1-\alpha$，即

$$P\left\{-Z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < Z_{\frac{\alpha}{2}}\right\} = 1-\alpha$$

从而，有 $P\left\{\overline{X} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$，即总体均值 μ 在 $1-\alpha$ 置信水平的置信区间为 $\left(\overline{X} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$。这里 $\frac{\sigma}{\sqrt{n}}$ 是抽样误差，$Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ 为一定倍数的抽样误差，称为极限误差范围，其意义是对给定的置信度进行区间估计所允许的最大误差。

（二）σ^2 未知时总体均值的区间估计

实际上，总体均值 μ 未知，而总体方差 σ^2 已知的情况是不常有的，通常的情况是二者都未知。设总体 $X \sim N(\mu,\sigma^2)$，σ^2 未知。此种情况下，可用样本标准差 S 来代替总体标准差 σ，根据 t 分布进行区间估计。为构造总体均值 μ 的置信区间，令 $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$，随机变量 T 服从自由度为 $n-1$ 的 t 分布，即 $T \sim t(n-1)$。由于 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$，其中 n 个偏差 $X_1 - \overline{X}, X_2 - \overline{X}, \cdots, X_n - \overline{X}$，有一个约束条件 $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$，即有 $n-1$ 个偏差可以自由取值，因此将 $n-1$ 称为自由度。给定置信度 $1-\alpha$，可以查 t 分布表得出自由度为 $n-1$ 的 t 分布的临界值 $t_{\frac{\alpha}{2},(n-1)}$，使得 $P\left\{-t_{\frac{\alpha}{2},(n-1)} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2},(n-1)}\right\} = 1-\alpha$，即

$$P\left\{\overline{X} - t_{\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}\right\} = 1-\alpha$$

当 $X \sim N(\mu,\sigma^2)$，且 σ^2 未知时，不能用 Z 统计量，只能以 T 统计量建立总体均值的置信区间，即 $\left(\bar{X}-t_{\frac{\alpha}{2},(n-1)}\dfrac{S}{\sqrt{n}},\ \bar{X}+t_{\frac{\alpha}{2},(n-1)}\dfrac{S}{\sqrt{n}}\right)$。该置信区间对偏离正态分布不太大的非正态总体也适用。

（三）两总体均值之差的区间估计

实际中，有时需要估计两个总体均值之间的差异，即需要对两个总体均值之差 $\mu_1-\mu_2$ 进行区间估计。限于篇幅，只给出两总体方差 ${\sigma_1}^2,{\sigma_2}^2$ 已知条件下它们均值之差的区间估计。由数理统计的知识，当 $X_1 \sim N(\mu_1,{\sigma_1}^2)$，$X_2 \sim N(\mu_2,{\sigma_2}^2)$ 时，样本均值之差 $\bar{X}_1-\bar{X}_2$ 也服从以 $\mu_1-\mu_2$ 为期望，$\dfrac{{\sigma_1}^2}{n_1}+\dfrac{{\sigma_2}^2}{n_2}$ 为方差的正态分布。此时，用于建立 $\mu_1-\mu_2$ 的置信区间的统计量为

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{{\sigma_1}^2}{n_1}+\dfrac{{\sigma_2}^2}{n_2}}}$$

它服从标准正态分布，可以利用正态分布表求得其数值。在给定显著水平 α 后，两个总体均值之差 $\mu_1-\mu_2$ 的置信区间可由下式确定：

$$\left[(\bar{X}_1-\bar{X}_2)-Z_{\frac{\alpha}{2}}\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}},\ (\bar{X}_1-\bar{X}_2)+Z_{\frac{\alpha}{2}}\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}\right]。$$

第三节 假设检验

假设检验是统计推断的另一类重要的问题。在总体的分布函数未知或只知其形式、但不知道其参数的情况下，为了推断总体的某些未知特性，提出某些关于总体的假设。例如，提出总体服从泊松分布的假设，又如，对于正态总体，提出数学期望等于 μ_0 的假设等。所作的假设是否成立，要根据样本提供的信息进行判断。它表

现为两个方面的问题：一是对事物的判断或估计是否准确；二是要通过一定统计方法来加以验证。利用样本的实际统计量，去检验事先对总体某些数量特征所作的假设，进而对假设做出接受还是拒绝的决策过程，就是假设检验。

假设检验作为一种以概率论和数理统计为基础的技术性方法，无论在自然科学领域还是在社会经济领域的统计分析中，都被广泛地采用。尤其是在经济学、生命科学和管理科学中，假设检验已成为一种必要的研究手段和工具。

一、假设检验的原理

假设检验首先是从对总体参数所作的一个假设开始，通过抽样获得的资料进行检验，进而判断对总体的认识是否正确可靠。假设命题一般包括两部分：原假设 H_0 和备择假设 H_1。若总体 X 中的参数 θ 未知，对 θ 提出“常规”的，或我们关心的并要对它正确与否做出判断的假设，称为原假设，和原假设对立的任何一个假设称为备择假设，分别记为 $H_0: \theta = \theta_0$ 和 $H_1: \theta \neq \theta_0$。原假设又称零假设，它往往是根据已有资料，或经过周密考虑后确定的。例如，要检验一批轴承是否符合平均直径为 10 毫米的标准，可以事先提出如下假设命题：该批轴承的平均直径为 10 毫米。然后从这批轴承中随机抽取样本，通过计算样本的平均直径来检验所作假设的正确性。用符号表示为 $H_0: \mu = 10$ mm。

假设检验中哪一个作为原假设，哪一个作为备择假设，要视具体的目的和要求而定。我们的目的是希望从样本观察值提供的信息对某陈述取得强有力的支持，则把这一陈述的否定作为原假设，而把陈述本身作为备择假设。一般来说原假设是受到“保护”的，没有足够的理由是不能随便拒绝原假设的。假设检验的目的就是要在原假设和备择假设之中选择一个，若认为原假设是正确的，则接受原假设，否则拒绝原假设而接受备择假设。为了充分理解假设

检验的原理，有必要介绍一下"显著性水平"这一概念。给出假设命题，相当于为总体参数建立了一个比较的标准。接下来的任务就是从总体中抽取样本，并根据实际观察的资料计算统计量的取值。通过样本统计量的取值与假设的总体参数比较来做出判断，一般情况下，两者完全一致的可能性是极小的，总会有差异存在，而且差异通常由两种情况引起：一种是由于工艺或试验条件的改变所引起的，称为条件差异；一种是生产或试验过程中偶然因素的影响所导致的结果的差异，称为随机差异。那么，差异达到什么程度才算显著呢？在上述两种原因所导致的差异中，如果样本统计量与假设总体参数间的差异超过了通常偶然因素起作用的程度，说明差异不仅仅是由随机因素引起的，据此可以否定总体的变动纯粹由随机因素引起，没有显著差异的原假设。换句话说，如果能证明统计量和假设的总体参数实际发生的差异超过给定的标准的可能性很小，那么，就有理由用反证法证明原假设是错误的，从而拒绝接受这个假设。当然，拒绝接受原假设，并不是因为它存在逻辑的矛盾，或实际上不存在这种假设，而仅仅因为它存的可能性很小。根据小概率原理，小概率事件在一次试验中几乎是不会发生的。如果根据原假设的条件计算出某一结果发生的概率很小，理应在一次试验中不至于发生，然而事实上又发生了，则有理由认为原假设是不成立的，从而拒绝接受。这里关键的问题是概率要小到什么程度才足以否定原来所作的假设。在进行假设检验时，应事先规定一个小概率标准，作为判断依据，这个小概率标准就称为显著性水平。由于原假设的分布是已知的，因而样本统计量和总体参数的离差在一定范围的概率可以求出，离差超过这个范围的概率当然也可以求出，如果统计量与参数差异过大，以至于发生这种事件的概率很小，而且小到低于给定的标准，我们自然就拒绝原假设；如果计算出的统计量与参数差异的相应概率大于给定的标准，我们就接受原假设。这样，把概率分布分为两个区间：离差的绝对值大于给

定的标准的概率分布区间称为拒绝区间，小于这个标准的则为接受区间。事件属于接受区间，原假设成立，判断总体无显著性差异；事件属于拒绝区间，则推翻原假设，认为总体有显著差异。

由于做出决策的依据是一个样本，当实际上 H_0 为真时，仍可能做出拒绝 H_0 的决策，这种以“真为假”的错误，在统计学中称为第一类错误，犯这种错误的概率记为：$P\{$拒绝 $H_0 \mid H_0$ 为真$\}$。我们无法排除犯这种错误的可能性，因此，自然希望将犯这类错误的概率控制在一定限度之内，即给出一个较小的数 $\alpha(0 < \alpha < 1)$，使犯这类错误的概率不超过 α，即 $P\{$拒绝 $H_0 \mid H_0$ 为真$\} \leqslant \alpha$。当原假设 H_0 不真时，依据样本做出接受 H_0 的判断，这种“以假为真”的错误通常称为第二类错误，犯第二类错误的概率记为：$P\{$接受 $H_0 \mid H_0$ 不真$\} = \beta$，其中 $\beta(0 < \beta < 1)$ 也是较小的正数。在确定检验法则时，我们希望犯两类错误的概率越小越好，但当样本容量 n 一定时，α 与 β 不可能同时减小，减小其中一个，另一个往往会增大，这说明承担风险是必然的。要它们同时减小，只有增加样本容量。实际中总是控制犯第一类错误的概率，通常取 $\alpha = 0.05$，0.005 等值。一般民意测验取 $\alpha = 0.01$，工程技术检验取 $\alpha = 0.001$ 甚至更低。

二、假设检验的步骤

下面通过一个例子，对假设检验的思路和程序进行归纳。

例 1　某工厂一直生产 60 W 标准灯泡，灯泡寿命近似服从正态分布 $N(100, 200^2)$，改进灯丝配料方案后，又生产一批灯泡，假设标准差 $\sigma = 200$ 不变。从新配料方案生产的一批灯泡中抽取 20 个，测试寿命值，得样本均值 $\bar{x} = 1\,100$，给定显著性水平 $\alpha = 0.05$，问灯泡的寿命是否有显著变化？[6]

解决上述问题，可以按以下思路和程序来进行。

(1) 提出原假设和备择假设

$$H_0: \mu = 1000, H_1: \mu \neq 1000$$

(2) 选取检验统计量

选取检验统计量的原则是必须含有参数的估计量,分布已知,并可以查表。该问题由于标准差已知,故选取服从正态分布的检验统计量为合适。当原假设为真时,有

$$Z = \frac{\bar{X} - 1000}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Z 不含任何未知参数,只要给定样本$(X_1, X_2, \cdots, X_n)$的任何一组观察值$(x_1, x_2, \cdots, x_n)$,就可以计算出 Z 的值,如果 Z 的值偏大到一定程度,就有理由拒绝原假设。

(3) 给定显著性水平 α, 求拒绝域

给定 $\alpha = 0.05$,使 $P(|Z| \geqslant z_{0.975}) = 0.05$。查表得分位数 $z_{0.975} = 1.96$,称 $z_{0.975}$ 为临界值,拒绝域为$(-\infty, -1.96]$或$[1.96, +\infty)$。

(4) 计算检验统计量的观察值

把 $\bar{x} = 1100, \sigma = 200, n = 20$ 代入 Z,求得 $z = 2.236$。

(5) 作判断

由于 $|z| > 1.96$,即 Z 的观察值落在拒绝域中,因此,可以认为改进灯丝配料方案后,灯泡的寿命有显著变化。如果要考察灯泡寿命是否有显著提高,则备择假设是单边的,即

$$H_0: \mu = 1000, H_1: \mu > 1000$$

给定 $\alpha = 0.05$,使 $P(Z \geqslant z_{0.975}) = 0.05$。查表得临界值 $z_{0.975} = 1.645$,拒绝域为$[1.645, +\infty)$。注意到拒绝域不等式的方向是沿着备择假设不等号的方向,且由于 $z > 1.645$,于是拒绝原假设,可以认为改进灯丝配料方案后,灯泡的寿命有显著提高。

三、总体参数检验

（一）总体均值检验

一个样本结果只是很多可能的结果中的一个。如果发现样本结果与标准值有差异，假设样本所在的总体均数与标准值相同，那么这种差异是否来自抽样，可以通过显著性检验来判定。根据样本含量的不同，两者之间的差异可以用不同的分布函数（如 t 分布，正态分布等）来描述。因此，可以利用这些分布来估计因抽样而导致的差异的可能性大小（即所谓的 α 水平），如果 α 很小，则有 $1-\alpha$ 的置信度，认为样本结果所代表的总体均数与标准值不同。统计学证明，样本均数 $\bar{X}$ 与总体参数（标准差 S）的差值与（$S/\sqrt{n}$）的商服从自由度为 $n-1$ 的 t 分布。因此，可以用 t 分布检验样本均数 $\bar{X}$ 与标准值 μ 间的差异归因于随机抽样的可能性。该方法称为 t 检验法，具体做法与上面例子中描述的方法类似。

（二）总体成数检验

表明现象的数据结构用成数（比率）指标。例如产品合格率、电视收视率、升学率、就业率等。要研究总体成数是否发生显著性变化，可以利用样本成数对总体成数进行假设检验。

总体成数可以看作总体（0,1）分布的平均数，其方差 $\sigma^2 = P(1-P)$，每个样本的（0,1）分布中，样本成数 p_i 也有 ${\sigma_i}^2 = p_i(1-p_i)$。由（0,1）分布和正态分布的关系，当样本容量 n 充分大的时候样本成数 p_i 趋近于一个平均数为 P，方差为 $\frac{P(1-P)}{n}$ 的正态分布，而统计量 $Z = \dfrac{p_i - P}{\sqrt{\dfrac{P(1-P)}{n}}}$ 趋近于标准正态分布。可以以该统计量，用 t 分布或 F 分布来进行假设检验。

（三）总体方差检验

方差反映现象在数量上的变异程度，或事物变化的均匀性程度。总体方差的假设检验方法可以用于检查总体方差是否会发生显著性变化。该检验常用于如下三种场合：① 当某一过程中，研究所关心的指标的变异性为质量标准，常常指定 σ^2 为某一规定值时；② 当已知其总体方差 σ^2，要用样本方差与之比较时；③ 当依据经验的理论来预测 σ^2 时。

1. 单个总体的假设检验

若总体 $X \sim N(\mu,\sigma^2)$，μ,σ^2 均未知。$X_1,X_2,\cdots,X_n$ 是来自 X 的样本。要求假设检验：$H_0:\sigma^2=\sigma_0{}^2$，$H_1:\sigma^2\neq\sigma_0{}^2$，其中 $\sigma_0{}^2$ 为已知常数。

由于 S^2 是 σ^2 的无偏估计，当 H_0 为真时，观察值 s^2 与 $\sigma_0{}^2$ 的比值一般接近于 1。此种情况下，$\dfrac{(n-1)S^2}{\sigma_0{}^2} \sim \chi^2(n-1)$，于是取 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0{}^2}$ 为检验统计量。拒绝域的形式为：$\dfrac{(n-1)s^2}{\sigma_0{}^2}\leqslant k_1$，或 $\dfrac{(n-1)s^2}{\sigma_0{}^2}\geqslant k_2$。$k_1,k_2$ 的值由下式确定：

$$P\{\text{拒绝 } H_0 \mid H_0 \text{ 为真}\} = P_{\sigma_0{}^2}\left\{\left(\frac{(n-1)S^2}{\sigma_0^2}\leqslant k_1\right)\cup\left(\frac{(n-1)S^2}{\sigma_0^2}\geqslant k_2\right)\right\}=\alpha$$

α 为给定的显著性水平。为方便计算，常取

$$P_{\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2}\leqslant k_1\right\}=\frac{\alpha}{2},\ P_{\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2}\geqslant k_2\right\}=\frac{\alpha}{2}$$

解得 $k_1=\chi^2_{1-\frac{\alpha}{2}}(n-1)$，$k_2=\chi^2_{\frac{\alpha}{2}}(n-1)$。于是拒绝域为

$$\frac{(n-1)s^2}{\sigma_0{}^2}\leqslant\chi^2_{1-\frac{\alpha}{2}}(n-1)，\text{或}\frac{(n-1)s^2}{\sigma_0{}^2}\geqslant\chi^2_{\frac{\alpha}{2}}(n-1)$$

这种检验方法称为 χ^2 检验法。同样的方法可以求单侧检验问题。

（1）右单侧检验：$H_0: \sigma^2 \leqslant {\sigma_0}^2, H_1: \sigma^2 > {\sigma_0}^2$。拒绝域为 $\chi^2 = \dfrac{(n-1)s^2}{{\sigma_0}^2} \geqslant \chi_\alpha^2(n-1)$。

（2）左单侧检验：$H_0: \sigma^2 \geqslant {\sigma_0}^2, H_1: \sigma^2 < {\sigma_0}^2$。拒绝域为 $\chi^2 = \dfrac{(n-1)s^2}{{\sigma_0}^2} \leqslant \chi_{1-\alpha}^2(n-1)$。

2. 两个总体的情况

设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别是来自正态总体 $N(\mu_1, {\sigma_2}^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的样本，且这两个样本相互独立。设样本方差分别为 ${S_1}^2$ 和 S_2^2，且 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均未知。在显著水平为 α 的前提下，假设检验为 $H_0: {\sigma_1}^2 \leqslant {\sigma_2}^2, H_1: {\sigma_1}^2 > {\sigma_2}^2$。

H_0 为真时，$E({S_1}^2) = {\sigma_1}^2 \leqslant {\sigma_2}^2 = E({S_2}^2)$；$H_1$ 为真时，$E({S_1}^2) = {\sigma_1}^2 > {\sigma_2}^2 = E({S_2}^2)$。当 H_1 为真时 $\dfrac{{S_1}^2}{{S_2}^2}$ 有偏大的趋势，故拒绝域有形式：$\dfrac{{s_1}^2}{{s_2}^2} \geqslant k$。常数 k 由下式确定：

$$P\{拒绝 H_0 \mid H_0 为真\} = P_{{\sigma_1}^2 \leqslant {\sigma_2}^2}\left\{\frac{{S_1}^2}{{S_2}^2} \geqslant k\right\} \leqslant P_{{\sigma_1}^2 \leqslant {\sigma_2}^2}\left\{\frac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2} \geqslant k\right\}$$

因为 ${\sigma_1}^2/{\sigma_2}^2 \leqslant 1$，要控制 $P\{拒绝 H_0 \mid H_0 为真\} \leqslant \alpha$，只要令

$$P_{{\sigma_1}^2 \leqslant {\sigma_2}^2}\left\{\frac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2} \geqslant k\right\} = \alpha$$

前已述及，$\dfrac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2} \sim F(n_1-1, n_2-1)$，于是由上式得，$k = F_\alpha(n_1-1, n_2-1)$。因此该检验问题的拒绝域为：$F = \dfrac{{s_1}^2}{{s_2}^2} \geqslant F_\alpha(n_1-1, n_2-1)$。

上述方法称为 F 检验法。

在总体参数的检验中，总希望所有真实的原假设都能得到接受，尽量避免真实的假设被拒绝，少犯或不犯第一类错误。也希望所有不真实的原假设都被拒绝，尽量避免不真实的假设被接受，少犯或不犯第二类错误。但由于两者是一对矛盾，在样本容量固定的情况下，两类错误不可能同时降低，因此在实际中，要根据两类错误可能造成的损失来做具体分析，由此决定尽可能减少哪一类错误。例如，对于实际无效的药物而决定大批量生产就会造成很大损失，因此要根据实际需要对显著性水平 α 进行控制，使其处于较低的水平，减少犯第一类错误的可能性；但把有显著效果的新药检验为无效果，以至于决定不投入生产，可能会导致某种疾病的蔓延，此种情况就要考虑降低犯第二类错误的可能性。更合理的做法是，比较两类错误可能造成损失的大小，然后再做决定。例如，对于上面例子，如果新药成本低廉，不妨冒犯第一类错误的风险；如果新药成本昂贵，则宁肯冒犯第二类错误的风险。

四、总体分布的拟合优度检验

拟合优度检验是用来判断一组样本观察值是否满足某种分布的一种假设检验方法。在参数的假设检验中，求检验的拒绝域时，需要利用已知的检验统计量的概率分布和给定的显著性水平。几乎每种情形，都假设样本是从正态总体抽取的，这使得所选取的检验统计量精确地服从正态分布、t 分布、χ^2 分布和 F 分布。参数的假设检验对样本的要求严格。一般地，样本取自正态总体，样本数据满足随机性、独立性与总体同分布性。然而，很多情况下，总体分布类型事先并不知道，为了实现某些参数的假设检验，必须首先对总体的分布类型进行假设检验，总体拟合优度检验即是利用样本观察值对总体的分布做出推断，从而判断总体是否服从特定的分布。其基本原理就是判断样本观察值与理论分布之间总的不符合

程度是否是由于抽样的偶然性造成的。由于实际中常用到的大都是离散型的随机变量，而且连续型随机变量的拟合优度检验原理过于复杂，下面只就离散型随机变量分布的拟合优度检验进行讨论。

二项分布、泊松分布、超几何分布都属于离散型变量分布，对于离散型随机变量分布的拟合优度检验比较适宜的方法是 χ^2 拟合优度检验法。

设一个随机试验的所有可能的结果有 k 个，设 k 个结果的对应值为 $1,2,3,\cdots,k$，令 p_i 表示出现第 i 个结果的概率。独立地做 n 次试验，每次试验都出现 k 个可能结果的一个，得到 n 个试验结果。实际上，是从某个多项分布的总体抽取容量为 n 的样本，这个多项分布对应的一个试验有 k 个可能的结果，令 $p_i = P(X = i), i = 1,2,\cdots,k, \sum_{i=1}^{k} p_i = 1$。检验假设为 $H_0: p_1 = p_{10}, p_2 = p_{20}, \cdots, p_k = p_{k0}$，$H_1: p_i \neq p_{i0}(i = 1,2,\cdots,k)$，对至少 i 的一个值。其中 $p_{i0}, i = 1,2,\cdots,k$ 是特定的多项分布中的参数，为已知。$0 < p_{i0} < 1, \sum_{i=1}^{k} p_{i0} = 1$。

令 N_i 表示 n 次试验中第 i 个结果出现的观察频数，np_{i0} 表示 H_0 为真时，第 i 个结果出现的期望频数或理论频数，$\sum_{i=1}^{k} N_i = n$，则 $(N_1, N_2, \cdots, N_k)$ 服从多项分布，其概率分布为

$$P(N_1 = n_1, N_2 = n_2, \cdots, N_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

简记为 $(N_1, N_2, \cdots, N_k) \sim M(n, p_1, p_2, \cdots, p_k)$，含有 k 个独立参数。为了用一个量来度量观察频数与期望频数之间的差异，考虑 $|N_i - np_i|$。当 H_0 为真时，$|N_i - np_{i0}|$ 很小；当 H_0 不真时，$|N_i - np_{i0}|$ 很大。Karl Pearson 于 1900 年提出如下形式的统计量：

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

并证明了当原假设 H_0 为真时，检验统计量 $\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_{i0})^2}{np_{i0}}$ 的极限分布是自由度为 $k-1$ 的 χ^2 分布。当样本容量足够大时，检验统计量 χ^2 近似服从 $\chi^2(k-1)$ 分布。在显著水平 α 下，使 $P(\chi^2 \geqslant \chi^2_{1-\alpha}(k-1)) = \alpha$。从 χ^2 分布表查得临界值 $\chi^2_{1-\alpha}(k-1)$，则得拒绝域为 $[\chi^2_{1-\alpha}(k-1), +\infty)$。如果 χ^2 的观察值落在拒绝域内，则拒绝原假设；否则接受原假设。χ^2 拟合优度检验法主要用来检验样本观察数据与理论分布的拟合优度。应用中需要注意以下几点：

(1) 检验统计量的极限分布是 χ^2 分布，因此，应抽取足够大的样本才能保证检验的效果。一般取 $n \geqslant 30$，或 $n \geqslant 50$，也可以用理论频数 np_{i0} 来调节，使之不小于5。

(2) χ^2 分布自由度的确定。当原假设中的理论分布不含任何未知参数，期望频数 np_{i0} 由计算直接得到，自由度为 $k-1$，因为有一个约束条件 $\sum_{i=1}^{k} N_i = n$。当原假设中的理论分布含有 r 个未知参数 $\theta_1, \theta_2, \cdots, \theta_r$，一般用极大似然估计量 $\hat{\theta}$ 代替 $\theta_i, 1, 2, \cdots, r$，，而 np_{i0} 的计算又取决于 $\hat{\theta}_i$，即取决于 r 个样本函数，因此自由度减少 r 个，χ^2 分布的自由度为 $k-1-r$。当自由度为1时，如果用 χ^2 统计量，需要做连续校正，校正公式为 $\chi^2 = \sum_{i=1}^{k} \frac{(|N_i - np_i| - 0.5)^2}{np_i}$。

(3) 当 k 较大时，检验统计量可以用等价形式 $\chi^2 = \sum_{i=1}^{k} \frac{N_i^2}{np_{i0}} - n$ 来代替，这样可以减少计算过程的误差。

下面通过几个例子说明 χ^2 拟合优度检验法的应用。

例2　根据遗传学规律，某种杂交花卉子代白花的比例为3∶1，在某时间种下该花种子，开花时节观察结果如表3.1实际频数行数据所示。试问开黄花和开白花的比例是否为3∶1？[1]

表 3.1　杂交花卉子代黄白花比例

频数	黄花	白花	合计
实际频数	84	16	100
理论频数	75	25	

解：H_0：黄花和白花的比例符合 3∶1 的理论比；H_1：黄花和白花的比例不符合 3∶1 理论比。

取 $\alpha = 0.05$

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_{i0})^2}{np_{i0}} = \frac{(84 - 75)^2}{75} + \frac{(16 - 25)}{25} = 4.320$$

自由度 $v = 1$，需做连续校正，故

$$\chi^2 = \sum_{i=1}^{k} \frac{(|N_i - np_i| - 0.5)^2}{np_i} = \frac{(|84 - 75| - 0.5)^2}{75} + \frac{(|16 - 25| - 0.5)^2}{25} = 3.853$$

查 χ^2 临界值表，得 $\chi^2_{0.05,1} = 3.841$，于是在 $\alpha = 0.05$ 的水平上拒绝 H_0，接受 H_1，认为黄花和白花比例不符合 3∶1 的理论比。

例 3　某工厂要对一批零件进行检验，检查了 100 个零件，结果如表 3.2 所示。问整批零件的疵点数是否服从泊松分布？

表 3.2　100 个零件的疵点数

点数	0	1	2	3	4	5	6
频数	14	27	26	20	7	3	3

解：设 H_0：疵点数服从泊松分布；H_1：疵点数不服从泊松分布。估计参数为 λ，由于只有一个被估参数，故 $r = 1$，此处分组数 $k = 7$，因此，自由度为 $k - r - 1 = 5$，$\chi^2_{0.05}(5) = 11.1$，在 H_0 成立的条件下，根据 $\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2(5)$，有 $\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = 3.55$，因为 $\chi^2 < \chi^2_{0.05}(5)$，所以接受原假设，即认为疵点数服从泊松分布。

对于其他离散型分布的拟合检验，方法相似，在此不作详述。

五、秩和检验

秩和检验是一种方便有效的检验方法，这种检验方法可用于下述的假设情形。

设有两个连续总体，它们的概率密度$f_1(x)$，$f_2(x)$均为未知，但已知$f_1(x)$，$f_2(x)$至多相差一平移，即$f_1(x) = f_2(x-a)$，a为未知常数。要检验下述各项假设：$H_0: a = 0, H_1: a < 0$；$H_0: a = 0, H_1: a > 0$；$H_0: a = 0, H_1: a \neq 0$。若两总体的均值$\mu_1$，$\mu_2$存在，则由于$f_1(x)$，$f_2(x)$至多相差一平移，因此有$\mu_2 = \mu_1 - a$。此时，上述各项假设分别等价于：$H_0: \mu_1 = \mu_2, H_1: \mu_1 < \mu_2$；$H_0: \mu_1 = \mu_2, H_1: \mu_1 > \mu_2$；$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$。

设X为一总体，将一容量为n的样本观察值按自小到大的顺序编号排列为

$$x_{(1)} < x_{(2)} < \cdots < x_{(n)}$$

称$x_{(i)}$的下标i为$x_{(i)}$的秩，$i = 1, 2, \cdots, n$。

设从两个总体中分别抽取容量为n_1，n_2的样本，且设两样本独立。这里总假定$n_1 < n_2$。将这$n_1 + n_2$个样本观察值放在一起，按自小到大的次序排列，求出每个观察值的秩，然后将属于第一个总体的样本观察值的秩相加，其和记为R_1，称为第一个样本的秩和，其余观察值的秩的总和记为R_2，称为第二个样本的秩和。显然R_1，R_2是离散型的随机变量，且有

$$R_1 + R_2 = \frac{1}{2}(n_1 + n_2)(n_1 + n_2 + 1)$$

因此，R_1，R_2中的一个确定后，另一个随之确定。我们只考虑统计量R_1即可。

考虑双侧检验问题$H_0: a = 0, H_1: a \neq 0$。直观上，当$H_0$为真时，则$f_1(x) = f_2(x)$。这时两个独立样本实际上来自同一总体，因

而第一个样本中各元素的秩应该随机地、分散地在自然数1 ~ $(n_1 + n_2)$ 中取值，一般来说不应过分集中取较小或较大的值。考虑到

$$\frac{1}{2}n_1(n_1 + n_2) \leqslant R_1 \leqslant \frac{1}{2}n_1(n_1 + 2n_2 + 1)$$

即知当 H_0 为真时秩和 R_1 一般不应取太靠近上面不等式两端的值。因而，当 R_1 的观察值 r_1 过大或过小时，都拒绝 H_0。因此，对于双侧检验问题 $H_0: a = 0, H_1: a \neq 0$，在给定显著性水平 α 下，H_0 的拒绝域为

$$r_1 \leqslant C_U\left(\frac{\alpha}{2}\right) \text{或} r_1 \geqslant C_L\left(\frac{\alpha}{2}\right)$$

其中临界点 $C_U\left(\frac{\alpha}{2}\right)$是满足 $P_{a=0}\left\{R_1 \leqslant C_U\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2}$ 的最大整数，而 $C_L\left(\frac{\alpha}{2}\right)$是满足 $P_{a=0}\left\{R_1 \geqslant C_L\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2}$ 的最小整数，它们可以通过查秩和临界值表来得到。而犯第一类错误的概率为

$$P_{a=0}\left\{R_1 \leqslant C_U\left(\frac{\alpha}{2}\right)\right\} + P_{a=0}\left\{R_1 \geqslant C_L\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

类似地，可得左侧检验 $H_0: a = 0, H_1: a < 0$ 的拒绝域为(显著性水平为 α) $r_1 \leqslant C_U(\alpha)$，此处，临界点 $C_U(\alpha)$ 是满足 $P_{a=0}\{R_1 \leqslant C_U(\alpha)\} \leqslant \alpha$ 的最大整数。而右侧检验问题 $H_0: a = 0, H_1: a > 0$ 的拒绝域为 $r_1 \geqslant C_L(\alpha)$，临界点 $C_L(\alpha)$ 是满足 $P_{a=0}\{R_1 \geqslant C_L(\alpha)\} \leqslant \alpha$ 的最小整数。

例 4　为查明某种血清是否会抑制白血病，选取患白血病已到晚期的老鼠9只，其中有5只接受这种治疗，另4只则不作治疗。从试验开始时计算其存活时间(单位:月)如下：

不作治疗	1.9	0.5	0.9	2.1	
接受治疗	3.1	5.3	1.4	4.6	2.8

设治疗与否的存活时间的概率密度至多相差一个平移，取 $\alpha = 0.05$，问这种血清对白血病是否有抑制作用？

解：本题实际上是检验接受治疗的老鼠的存活期是否有延长。在给定的条件下，两组样本可以看成是独立的。设不作治疗和接受治疗的老鼠的存活期总体的均值分别为 μ_1,μ_2。于是，所作的假设为 $H_0:\mu_1=\mu_2, H_1:\mu_1<\mu_2$。这里，$n_1=4, n_2=5, \alpha=0.05$。先计算对应于 $n_1=4$ 的一组观察值的秩和。将两组数据放在一起按自小到大的顺序排列。对来自第一个总体（$n_1=4$）的数据下面加一下划线表示，即有

数据	$\underline{0.5}$	$\underline{0.9}$	1.4	$\underline{1.9}$	$\underline{2.1}$	2.8	3.1	4.6	5.3
秩	1	2	3	4	5	6	7	8	9

所以 R_1 的观察值为 $r_1=1+2+4+5=12$。查秩和临界值表得 $C_U(0.05)=12$，即拒绝域为 $r_1\leqslant 12$，故拒绝 H_0，即认为这种血清对白血病有抑制作用。

六、独立性检验

从同一总体中随机抽取若干个样本，计算第一个样本率，所得的样本率有大有小。这种差异是由于抽样误差所引起的，因此在比较两个样本率或多个样本率时，不能仅从样本率的大小判断它们所代表的总体率是否有差异，用样本率比较反映总体率的差异，必须考虑因抽样引起的变异，即必须进行假设检验。

（一）四格表资料的 χ^2 检验

下面介绍一下医学论文中常用的四格表资料的 χ^2 检验。

什么是四格表资料？一般情况下，凡是两个率或构成比资料都可以看做四格表资料。例如，表 3.3 就可以看作是一个四格表资料。

表 3.3　四格表资料示例表

组别	发病人数	未发病人数	观察例数	发病率/%
实验组	14	86	100	14
对照组	30	90	120	25
合计	44	176	220	20