数理统计简介

- 数理统计
  - 采集样本
    - 抽样技术
      - 怎样抽
      - 抽多少
    - 试验设计
  - 统计推断
    - 统计估计
      - 参数估计
        - 点估计
        - 区间估计
      - 非参数估计
    - 统计检验
      - 参数假设检验
      - 非参数假设检验
  - 统计分析
    - 方差分析
    - 回归分析

第七章　样本分布

§7.1　总体与样本

§7.3　样本分布的数字特征

§7.4　几个常用统计量的分布

7.1 总体与样本

定义7.1. 总体：研究对象的全体。

通常指研究对象的某项数量指标。

组成总体的每个基本单位称为**个体。**

从本质上讲，总体就是所研究的随机变量或随机变量的分布。

总体可以包含有限个个体, 也可以包含无限个个体. 在一个有限总体所包含的个体相当多的情况下, 可以把它作为无限总体来处理. 例如, 一麻袋稻谷, 一个国家的人口.

每一总体中的个体,具有共同的可观察的特征,把它作为不同总体的区别.

例如,灯泡厂一天生产5万个25万瓦白炽灯泡,按规定,使用寿命不足0.1万小时的为次品.在考察这批灯泡的质量时,“该天生产的5万个25瓦白炽灯泡的全体”组成一个总体,每一个灯泡是总体中的一个个体,其共同的可观察的特征为灯泡的使用寿命.

数轴上的“一条线段所有点的全体”组成一个总体,其中的每一个点是总体的一个个体,其共同的可观察的特征为点在数轴上的位置.

对于一个总体来说,其每一数量特征就是一个随机变量ξ .由于人们主要是研究总体的某些数量特征,所以把总体看作所研究对象的若干数量特征的全体,而直接用一个随机变量ξ(也可以是一个多元随机变量)的代表.

定义7.2 总体中抽出若干个体而成的集体,称**样本**.样本中所含个体的个数,称为**样本容量**。

来自总体的样本X_1, … , X_n **满足:**

(1)**代表性:**X_1,…,X_n与总体服从相同的分布.

(2)**独立性:**X_1, ..., X_n相互独立;

在进行抽样时, 样本的选取必须是随机的, 即总体中每个个体都有同等机会被选入样本. 抽样通常有两种方式：一种是**不重复抽样**, 即每次抽取一个不放回去, 再抽取第二个, 连续抽取n次；另一种是**重复抽样**, 指每次抽取一个, 进行观察后再放回去, 再抽取第二个, 连续抽取n次, 构成一个容量为n的样本.

简单随机样本：进行重复抽样所得的随机样本称为简单随机样本.

如上所述，所谓总体就是一个随机变量，所谓样本就是n个相互独立且与总体有相同分布的随机变量 $X_1, \cdots, X_n$（n是样本容量）. 通常把它们看成一个n元随机变量$(X_1, \cdots, X_n)$，而每一次具体抽样所得的数据，就是n元随机变量的一个观察值(样本值)，记为$(x_1, \cdots, x_n)$.

一个容量为n的样本有双重意义：有时指一次抽样的具体数值$(x_1, \cdots, x_n)$，有时泛指一次抽出的可能结果，这就是指一个n元随机变量. 用大写字母$(X_1, \cdots, X_n)$表示.

定义7.3：称样本X_1， …，X_n的函数$f(X_1, \cdots, X_n)$是总体X的一个**统计量**，**如果$f(X_1, \cdots, X_n)$不含未知参数**。

§7.2 样本分布函数

设 $x_1, x_2, \ldots, x_n$ 是取自总体的样本，若将样本观测值由小到大进行排列, 为 $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$， 则称 $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ 为有序样本,

用有序样本定义如下函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ k/n, & x_{(k)} \le x < x_{(k+1)}, \quad k=1,2,\ldots,n-1 \\ 1, & x_{(n)} \le x \end{cases}$$

§7.3样本分布的数字特征

样本的数字特征, 是显示一个样本分布某些特征的数字. 人们经常用它们来估计总体的数字特征.

(一)样本平均数

定义7.4 对于样本($X_1, \ldots, X_n$), 称

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (7.1)$$

为样本平均数.

- 对于某具体样本值(x_1， … x_n)，样本平均数是

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(二)样本方差

定义7.5　对于样本$(X_1 \cdots X_n)$,称

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (7.3)$$

以及　$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

分别为样本方差和样本标准差.

由(7. 3)式, 有 $$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i \overline{X} + n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2\right)$$

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right) \qquad (7.4)$$

§7.4几个常用统计量的分布

后面的几章,所涉及的多为正态分布,因此,这一节里将介绍有关正态分布随机变量函数的一系列分布.其中有些定理(如定理7.3-定理7.5)的证明用到较多的线性代数知识,书中没有进行证明.对于定理7.1和定理7.2也只是强调它们的结论本身.

定理 7.1 设$X_1,X_2,\ldots,X_n$相互独立, $X_i\sim N(\mu_i, \sigma_i)$, $i=1,2,\ldots,n$, 则它们的线性函数

$$\eta=\sum_{i=1}^{n}a_iX_i\,(a_i\text{不全为零}),\text{也服从正态分布},$$

$$\eta\sim N\left(\sum_{i=1}^{n}a_i\mu_i,\sum_{i=1}^{n}a_i^2\sigma_i^2\right)$$

独立的服从正态分布的随机变量的
非零线性组合仍服从正态分布

推论 设$(X_1,X_2,\ldots,X_n)$是取自正态总体$N(\mu,\sigma^2)$的样本, 则有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

则 $(1)\ \overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$

$$(2)\ \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

定理 7.2 设$X_1,X_2,\ldots,X_n$相互独立, $X_i \sim N(0, 1)$, $i=1,2,\ldots,n$, 则

$$X_1^2 + X_2^2 + \cdots + X_n^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

即n个相互独立的标准正态分布的随机变量的平方和服从n个自由度的$\chi^2(n)$分布

定理 7.3设$X_1,X_2,\ldots,X_n$相互独立, $X_i \sim N(0, 1)$, $i=1,2,\ldots,n$, 则

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\eta = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{i=1}^{n} X_i \right)^2$$

则η与$\overline{X}$相互独立,

$$\eta \sim \chi^2(n-1)$$

推论　设$(X_1,X_2,\ldots,X_n)$是取自正态总体$N(\mu,\sigma^2)$的样本, 则有

$$(1)\ \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2(n-1)$$

$$\text{即}\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(2) $\overline{X}$与$\sum_{i=1}^{n}(X_i-\overline{X})^2$相互独立

即$\overline{X}$与S^2相互独立

定理7.4 设两个随机变量ξ与η相互独立, 并且$\xi \sim N(0,1)$, $\eta \sim \chi^2(n)$, 则

$$T = \frac{\xi}{\sqrt{\eta / n}}$$服从具有n个自由度的t分布

或记作$T = \dfrac{\xi}{\sqrt{\eta / n}} \sim t(n)$

推论1 设$(X_1,X_2,\ldots,X_n)$是取自正态总体$N(\mu,\sigma^2)$的样本,

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

则

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$$

定理 7.5 设两个随机变量ξ_1和ξ_2相互独立, 且 $\xi_1 \sim \chi^2(n_1), \xi_2 \sim \chi^2(n_2)$, 则有

$$F = \frac{\xi_1 / n_1}{\xi_2 / n_2} \sim F(n_1, n_2)$$

其中$F(n_1, n_2)$为第一个自由度是n_1,第二个自由度是n_2的F分布.

推论 设设$X_1,X_2,\ldots,X_n$和$Y_1,Y_2,\ldots,Y_m$分别来自两个相互独立的正态总体

$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$, 则

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

其中, S_1^2,S_2^2分别为两个样本各自的方差.