第6.2节　样本数字特征

一、统计量

二、样本均值

三、样本方差与样本均方差

四、样本矩

五、小结

一、统计量

由样本推断总体特征,需要对样本值进行“加工”,“提炼”.这就需要构造一些样本的函数,它把样本中所含的信息集中起来.

定义1 设$X_1,X_2,\ldots,X_n$是来自总体X的一个样本, $g(X_1,X_2,\ldots,X_n)$是$X_1,X_2,\ldots,X_n$的函数, 若g中不含未知参数, 则$g(X_1,X_2,\ldots,X_n)$称是一统计量.

因为$X_1,X_2,\ldots,X_n$都是随机变量, 而统计量$g(X_1,X_2,\ldots,X_n)$是随机变量的函数, 因此统计量是一个随机变量. 我们将统计量的分布称为抽样分布.设是$x_1,x_2,\ldots,x_n$相应于样本的样本值, 则称$g(x_1,x_2,\ldots,x_n)$是$g(X_1,X_2,\ldots,X_n)$的观察值.

例1 设 X_1, X_2, X_3 是来自总体 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 为已知，σ^2 为未知，判断下列各式哪些是统计量，哪些不是？

$$T_1 = X_1, \qquad T_2 = X_1 + X_2 e^{X_3},$$

$$T_3 = \frac{1}{3}(X_1 + X_2 + X_3), \qquad T_5 = X_1 + X_2 - 2\mu,$$

$$T_4 = \max(X_1, X_2, X_3),$$

是

$$T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2).$$

不是

二、 样本均值

定义2 设$X_1,X_2,...,X_n$是来自总体X的一个样本, 令

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

则称 $\overline{X}$ 为样本均值或样本平均数 .

对应的观察值： $\overline{x} = \frac{1}{n}\sum_{i=}^{n} x_i$

注：样本均值为样本的平均数，用来表征观察值的“中心”，反映了总体一般水平的综合指标.

定理1 设$X_1,X_2,\ldots,X_n$是来自总体X的一个样本, 若$E(X)=\mu,D(X)=\sigma^2$，则

(1)$E\left(\overline{X}\right)=\mu, D\left(\overline{X}\right)=\dfrac{\sigma^2}{n}$;

(2)当n充分大时,$\overline{X}$近似服从正态分布$N\left(\mu,\sigma^2\right)$.

三、样本方差与样本均方差

定义**3** 设$X_1, X_2, \ldots, X_n$是来自总体X的一个样本, 令

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

则称 S^2 为样本方差，且称 S^2 的算术根

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

为样本均方差或样本标准差 .

注：样本方差和样本均方差都是统计量，样本方差反映了样本数据的离散程度。

定理**2** 设$X_1,X_2,\ldots,X_n$是来自总体X的一个样本，若$D(X)=\sigma^2$，则

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

$$(2)E\left(S^2\right)=\sigma^2.$$

四、 样本矩

定义**4** 设$X_1,X_2,...,X_n$是来自总体X的一个样本**,** 则统计量

$$a_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,k=1,2,\cdots$$

称为样本**k**阶原点矩，统计量

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k,\ k=2,3,\cdots$$

称为样本**k**阶中心距.

注：样本一阶原点矩为样本均值；

样本二阶中心距不是样本方差。

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

$$E\left(S_n^2\right)=\frac{n-1}{n}\sigma^2$$

样本方差及样本均方差的观察值分别为

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right);$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

样本k阶原点矩及样本k阶中心距的观察值分别为

$$\alpha_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k,\ \ k = 1,2,\cdots;$$

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k,\quad k = 2,3,\cdots.$$

这些观察值仍分别称为样本方差, 样本标准差, 样本 k阶(原点)矩以及样本k阶中心矩.

辛钦定理

设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立，服从同一分布，且具有数学期望$E(X_k) = \mu \ (k = 1,2,\cdots)$,则对于任意正数$\varepsilon$,有$\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$

费歇资料

Ronald Aylmer Fisher

Born: **17 Feb 1890 in London, England**

Died: **29 July 1962 in Adelaide, Australia**

学生氏资料

William Sealey Gosset

Born: 13 June 1876 in Canterbury, England

Died: 16 Oct 1937 in Beaconsfield, England

格里汶科资料

Boris Vladimirovich Gnedenko

Born: 1 Jan 1912 in Simbirsk (now Ulyanovskaya), Russia

Died: 27 Dec 1995 in Moscow, Russia