# 统计量

数理统计的基本任务之一是利用样本所提供的信息来对总体分布中未知的量进行推断，但是样本观测值常常表现为一大堆数字，很难直接用来解决我们所要研究的具体问题．人们常常把数据加工成若干个简单明了的数字特征，由数据加工后的数字特征就是统计量的观测值．

**定义6.2.1** 不含有未知参数的样本的函数 $g\left(X_1,\cdots,X_n\right)$ 称为统计量.

例1　假设总体 $X \sim N\left(\mu,\sigma^2\right),\left(X_1,X_2,\cdots,X_n\right)$为取自该总体的一个样本，其中$\mu$未知而$\sigma^2$已知．设$n=4$，试判断下列表达式是否为统计量：

$$(1)\ \sum_{i=1}^{4} X_i^2;\quad (2)\ \frac{1}{4}\sum_{i=1}^{4}\left(X_i-\mu\right)^2;\quad (3)\ \frac{1}{\sigma^2}\sum_{i=1}^{4}\left(X_i-\bar{X}\right)^2.$$

解　由定义即知(2)不是统计量，而(1)(3)是．

1.常用统计量介绍

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

样本均值的观测值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

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

样本方差的观测值 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

样本标准差 $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}$

样本标准差的观测值 $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$

样本的二阶中心矩 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$

相应的观测值 $s_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$

注 $S^2, S_n^2$在计算时的另一表达形式:

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

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}\cdot n\bar{X}+n\bar{X}^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$$

由此知

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

显然有 $$(n-1)S^2 = nS_n^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

(3)样本的 $k$ 阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad \left(A_1 = \bar{X}\right)$

(4)样本的 $k$ 阶中心矩 $M_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k \quad \left(M_2 = S_n^2\right)$

其中 $k$ 是正整数，而每个统计量也都有相应的观测值.

(5)次序统计量

设$(X_1, X_2, \cdots, X_n)$为取自总体 $X$ 的一个样本，其中取值最大的记为$X_{(n)}$，称为最大次序统计量；而取值最小的记为$X_{(1)}$，称为最小次序统计量；$X_{(1)}$与$X_{(n)}$统称为次序统计量.

# 常用统计量的性质

**定理** 设总体$X$的均值$E(X)=\mu$，方差$D(X)=\sigma^2$; $(X_1, X_2, \cdots, X_n)$为取自该总体的一个样本，则

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

(2) $E(S^2)=\sigma^2, E(S_n^2)=\dfrac{n-1}{n}\sigma^2, n \geq 2$.

证(2)

$$(n-1)E\left(S^2\right)=E\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)=\left(\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\bar{X}^2\right)\right)$$

$$=n\left(D\left(X_1\right)+E^2\left(X_1\right)-D\left(\bar{X}\right)-E^2\left(\bar{X}\right)\right)$$

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

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

例2　设总体 $X \sim R(-1,3)$, $(X_1, X_2, \cdots, X_{10})$ 为来自总体的一个样本，求 $E(\bar{X}), D(\bar{X}), E(S^2)$.

解　由题设条件知

$$E(X) = \mu = 1, D(X) = \sigma^2 = \frac{4}{3}, n = 10$$

所以由定理可得

$$E(\bar{X}) = \mu = 1, D(\bar{X}) = \frac{\sigma^2}{n} = \frac{2}{15}, E(S^2) = \frac{4}{3}.$$

2.次序统计量的分布

设总体 $X$ 具有分布函数 $F(x)$ 及概率密度函数 $f(x)$,

则最小次序统计量具有概率密度函数:

$$f_1(y)=n\left[1-F(y)\right]^{n-1}f(y)$$

最大次序统计量具有概率密度函数:

$$f_n(y)=nF(y)^{n-1}f(y).$$

由第三章中关于独立同分布情形下求最大或最小随机变量的分布函数之结论知

$$F_n(y)=F^n(y), F_1(y)=1-\left(1-F(y)\right)^n$$

各自求导即得.

# 谢　谢

同济大学数学科学学院概率统计教学团队