概率论与数理统计

第四节 统计量及抽样分布

一、 统计量

二、 抽样分布

一　统计量

1、定义 设$X_1, X_2, \cdots, X_n$是总体X的容量为n的样本，$T(X_1, X_2, \cdots, X_n)$是定义在样本空间上不依赖于未知参数的连续函数，则称$T(X_1, X_2, \cdots, X_n)$为统计量

2、一点说明

统计量$T(X_1, X_2, \cdots, X_n)$为随机变量，当样本$X_1, X_2, \cdots, X_n$取定观测值$x_1, x_2, \cdots, x_n$时，$T(x_1, x_2, \cdots, x_n)$为常量或观测值。

第十六次课结束

例1 设$X_1,X_2,\cdots,X_n$来自正态总体$N(\mu,\sigma^2)$的样本,其中μ已知,σ^2未知,问下面哪些是统计量?

(1) $\frac{1}{3}\sum_{i=1}^{n}X_i$ 是

(2) $\frac{1}{\sigma^2}\sum_{i=1}^{4}(X_i-\overline{X})^2$,其中$\overline{X}=\frac{1}{4}\sum_{i=1}^{4}X_i$ 否

(3) $\sum_{i=1}^{5}X_i^2$ 是

(4) $\sum_{i=1}^{5}(X_i-\mu)^2$ 是

(5) $\max\{X_1,X_2,\cdots,X_n\}$ 是

3、几个常用的统计量

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

它反映了总体均值的信息

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

它反映了总体方差的信息

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

（3）样本k阶原点矩：$a_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$

（4）样本k阶中心矩：$A_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$

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

（5）顺序统计量:

若$x_1, x_2, \cdots, x_n$ 是样本观测值，将其按照从小到大的顺序重新排列$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

定义$X_{(i)}$，使得不论样本$X_1, X_2, \cdots, X_n$取怎样的观察值 $x_1, x_2, \cdots, x_n$，$X_{(i)}$总是以$x_{(i)}$为观察值$(i = 1, 2, \ldots, n)$，

显然$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，$X_{(i)}$称为第$i$ 顺序统计量

$$X_{(1)} = \min_{1 \le i \le n} X_i，为最小顺序统计量$$

$$X_{(n)} = \max_{1 \le i \le n} X_i，为最大顺序统计量$$

（6）样本中位数：

$$M=\begin{cases} X_{(\frac{n+1}{2})} & n\text{为奇数} \\ \frac{1}{2}\left(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}\right) & n\text{为偶数} \end{cases}$$

（7）样本极差：

$$R=X_{(n)}-X_{(1)}$$

4、性质

设$X_1, X_2, \cdots, X_n$是取自总体X 的简单随机样本,

$EX=\mu, DX=\sigma^2$,则有

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

(2) $ES^2=\sigma^2, ES^{*2}=\dfrac{n-1}{n}\sigma^2.$

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

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

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

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

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

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

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

$$E[\sum_{i=1}^{n}(X_i-\bar{X})^2]=\mathrm{E}[\sum_{i=1}^{n}X_i{}^2-n\bar{X}^2]$$

$$\mathrm{E}[\sum_{i=1}^{n}(X_i-\bar{X})^2]=\mathrm{E}[\sum_{i=1}^{n}X_i^{\,2}-n\bar{X}^{\,2}]$$

$$=\sum_{i=1}^{n}\mathrm{E}X_i^{\,2}-n\mathrm{E}\bar{X}^{\,2}$$

$$=n(\sigma^2+\mu^2)-n(\frac{1}{n}\sigma^2+\mu^2)=(n-1)\sigma^2$$

$$\mathrm{E}X_i^{\,2}=\mathrm{D}X_i+(\mathrm{E}X_i)^2=\sigma^2+\mu^2$$

$$\mathrm{E}\bar{X}^{\,2}=\mathrm{D}\bar{X}+(\mathrm{E}\bar{X})^2=\frac{1}{n}\sigma^2+\mu^2$$

$$\mathrm{E}S^2=\mathrm{E}[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2]=\sigma^2$$

设$X_1,X_2,\cdots,X_n$是取自总体 为$\chi^2(n)$分布的简单随机样本,
则$E\overline{X}=($　　　$)$, $D\overline{X}=($　　　$)$.

设总体$X\sim N(\mu_1,\sigma^2)$,$Y\sim N(\mu_2,\sigma^2)$, X与Y相互独立,
$X_1,X_2,\cdots,X_{n_1}$取自总体X 的简单随机样本,$Y_1,Y_2,\cdots,Y_{n_2}$为
取自总体Y 的简单随机样本,则

$$E\left[\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2}{n_1+n_2-2}\right]=(\quad\quad)$$

二　抽样分布

1　单个正态总体的情况

设总体$X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_n$是取自总体X的简单随机样本,且样本均值$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$,样本方差

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$ 样本二阶中心距$$S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

则有 (1)$\bar{X} \sim N(\mu,\frac{\sigma^2}{n})$

(2)$\bar{X}$与S^2相互独立;

$$(3)\ \frac{n-1}{\sigma^2}S^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{n}{\sigma^2}S^{*2} \sim \chi^2(n-1)$$

推论 (1) $\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$

(2) $\frac{\bar{X}-\mu}{S}\sqrt{n}\sim t(n-1)$ 或 $\frac{\bar{X}-\mu}{S^{*}}\sqrt{n-1}\sim t(n-1)$

---

(1) $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$

(2) $\bar{X}$与S^2相互独立；

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

2 两个正态总体的情况

设总体$X\sim N(\mu_1,\sigma_1^2)$,$Y\sim N(\mu_2,\sigma_2^2)$, X与Y相互独立, $X_1,X_2,\cdots,X_{n_1}$ 取自总体X 的简单随机样本, $Y_1,Y_2,\cdots,Y_{n_2}$为取自总体Y 的简单随机样本,记

$$\overline{X}=\frac{1}{n_1}\sum_{i=1}^{n_1}X_i \qquad S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2$$

$$\overline{Y}=\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i \qquad S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2$$

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

(1) 在上述定理条件下, 若$\sigma_1^2=\sigma_2^2=\sigma^2$, 则

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

(2) 在上述定理条件下, 若$\sigma_1^2\neq\sigma_2^2$, 则

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

一、求统计量的数字特征

例1 设总体$X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_{2n}$是取自总体X的简单随机样本, 且样本均值$\bar{X} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$, 样本方差

求统计量$Y = \sum_{i=1}^{n}(X_i + X_{n+i} - 2\bar{X})^2$ 的数学期望EY

二、求统计量的分布

例2 设$X_1, X_2, \cdots, X_9$ 是来自总体X 的简单随机样本,

令$Y_1 = \frac{1}{6}\left(X_1 + X_2 + \ldots X_6\right), Y_2 = \frac{1}{3}\left(X_7 + X_8 + X_9\right)$

$S^2 = \frac{1}{2}\sum_{i=7}^{9}\left(X_i - Y_2\right)^2; Z = \frac{\sqrt{2}\left(Y_1 - Y_2\right)}{S}$

求统计量Z 的分布

例3 $X \sim N(0,3^2), Y \sim N(0,3^2), X_1, X_2, \cdots, X_9$来自总体$X$,

$Y_1, Y_2, \ldots, Y_9$来自总体Y的简单$r.v$ 样本,

则统计量$U = \dfrac{X_1 + X_2 + \ldots + X_9}{\sqrt{Y_1^2 + \ldots + Y_9^2}} \sim$

三、求统计量取值的概率

例4 从正态总体$N(\mu, 0.5^2)$中抽取样本$X_1, X_2, \cdots, X_{10}$

(1)已知$\mu=0$,求$P\left\{\sum_{i=1}^{10} X_i^2 \geq 4\right\}$

(2)μ未知,求$P\left\{\sum_{i=1}^{10}(X_i - \overline{X})^2 \geq 2.85\right\}$

例5 $X_1, X_2, \cdots, X_{26}$为$N\left(0, \sigma^2\right)$的样本,

$$\text{求}P\left\{\frac{\sum_{i=1}^{10} X_i}{\sqrt{\sum_{j=11}^{26} X_j^2}} \leq 1.6795\right\}$$

例6 $X_1,X_2,\cdots,X_n$取自总体$X\sim N(\mu,\sigma^2)$的简单随机样本，$\overline{X}$是样本均值，记

$$S_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2;\quad S_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2;$$

$$S_3^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2;\quad S_4^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

则服从自由度为$n-1$的t分布的统计量为

(A) $\dfrac{\overline{X}-\mu}{S_1}\sqrt{n-1}$; (B) $\dfrac{\overline{X}-\mu}{S_2}\sqrt{n-1}$;

(C) $\dfrac{\overline{X}-\mu}{S_3}\sqrt{n}$; (D) $\dfrac{\overline{X}-\mu}{S_4}\sqrt{n}$.