第五章　数理统计中的统计量及其分布

内容提要

（一）随机样本和经验分布函数

1. 总体

常把研究对象 $X(\omega)$ 称为总体（或母体）。

2．样本

把对母体进行 n 次采样以后所得到的结果 $(X_1, X_2, \cdots X_n)$ 称之为样本。

3．简单随机样本

由于重复采样时结果会不同，故样本亦为随机的，通常假设 X_i 之间相互独立，且它们都服从与母体 X 相同的分布，对于这种独立同分布的随机变量，称之为简单随机样本，常简记作 i．i．d.

4．样本观察值

采样次数 n 称为样本容量，而样本的一组具体值称为样本观察值，记为 $(x_1, x_2, \cdots, x_n)$。

5．经验分布函数

将样本的观察值 $x_1, x_2, \cdots x_n$ 按大小重新排列为 $x_{(1)}, x_{(2)}, \cdots x_{(n)}$，则有 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，定义函数

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

为样本的经验分布函数，记作 $F_n(x)$。

6. 分位数

当随机变量 $T \sim t(n)$ 时，称满足关系式：

$$P(T \le t_{1-\alpha}(n)) = 1 - \alpha,$$

的数 $t_{1-\alpha}(n)$ 为自由度为 n 的 t 分布的“$1-\alpha$ 临界值”（Critical Value）或“$1-\alpha$ 分位数”。

（二）常用统计量

1. 样本均值 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$，

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

3. 样本标准差：

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

4．样本k阶（原点）矩

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

5.次序统计量

将样本的观察值 $x_1,x_2,\cdots x_n$ 按大小重新排列为 $x_{(1)},x_{(2)},\cdots x_{(n)}$ ，则有 $x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$，称 $X_{(i)}$ 为样本的第 i 个次序统计量.

6.中位数Me

$$Me=\begin{cases} x_{(\frac{n+1}{2})}, & n\text{为奇数} \\ \frac{1}{2}[x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}], & n\text{为偶数。}\end{cases}$$

7. 极差R

$$R=\max X_i-\min X_i=x_{(n)}-x_{(1)}$$

(三) 抽样分布

1. χ^2−分布

设 $X_1,X_2,\cdots,X_n$ 独立同分布于标准正态分布 $N(0,1)$，令 $\chi^2\overset{\Delta}{=}{X_1}^2+{X_2}^2+\cdots+{X_n}^2$，则称 χ^2 服从的分布为“自由度”为 n 的 χ^2 分布，记为 $\chi^2\sim\chi^2(n)$。

$\chi^2(n)$ 分布具有如下几个性质：

(1) 若 $X\sim\chi^2(n)$， $EX=n,\ DX=2n$。

(2) $\chi^2(n)$ 分布具有可加性，即若 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 与 Y 独立，则 $X+Y\sim\chi^2(m+n)$。

（3） $\chi^2(n)$ 分布不是对称分布。是 n 个独立的 $N(0,1)$ 的平方和，其取值范围为 $[0,+\infty)$.

2. F −分布

设 $X\sim\chi^2(m),Y\sim\chi^2(n)$，且 X 与 Y 相互独立，令 $F\overset{\Delta}{=}\dfrac{X/m}{Y/n}$，则称随机变量 F 服从的分布为“自由度”为 m与n 的 F 分布，记为 $F\sim F(m,n)$.

F -分布性质：

(1) F 分布也是一个只取非负值的偏态分布;

(2) $F_\alpha(n,m)=\dfrac{1}{F_{1-\alpha}(m,n)}$。

3. t 分布

设随机变量 X 与 Y 相互独立，且 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，则称 $T\overset{\Delta}{=}\dfrac{X}{\sqrt{Y/n}}$ 服从的分布为“自由度”为 n 的 t 分布，记为 $T\sim t(n)$。

t 分布性质：

(1) 密度函数 $\varphi(x)$ 关于纵轴 $x=0$ 对称。

(2) $n>1$ 时，t 分布的数学期望存在，且为 0；

$n>2$ 时，t 分布的方差存在，且为 $n/(n-2)$。

(3) 与 χ^2 分布类似，自由度 n 的大小也具有某种信息量的含义。从量纲角度看，是无量纲的。

（四）正态总体下常用统计量的一些重要结论

1.单正态总体的常用统计量的分布

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

（1） $\overline{X}\sim N(\mu,\dfrac{\sigma^2}{n})$；

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

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

(4) $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，$(\overline{X},S^2)$ 相互独立的;

(5) $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\dfrac{(n-1)S^2+n(\overline{X}-\mu)^2}{\sigma^2}\sim\chi^2(n)$；

2.两个正态总体的常用统计量的分布

$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$，其样本容量分别为$m,n$的，它们的样本相互独立，

$\overline{X},\overline{Y}$分别表示其样本均值，则有

（1）$U=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1{}^2}{m}+\dfrac{\sigma_2{}^2}{n}}}\sim N(0,1)$；

（2）S_X^2,S_Y^2分别表示其样本方差，当$\sigma_1=\sigma_2$则有

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}\sim t(m+n-2)$$

其中$S_W=\sqrt{\dfrac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}$，即$S_W{}^2$可视为样本方差$S_X^2$与$S_Y^2$的加权平均值。

（3）$\dfrac{S_x{}^2\big/\sigma_1{}^2}{S_y{}^2\big/\sigma_2{}^2}\sim F(m-1,n-1)$；

（4）$\dfrac{[(m-1)S_x{}^2/m+(\overline{X}-\mu_1)^2]\big/\sigma_1{}^2}{[(n-1)S_y{}^2/n+(\overline{Y}-\mu_2)^2]\big/\sigma_2{}^2}\sim F(m,n)$；