哈爾濱工業大學

统计量及抽样分布

统计量

由样本值去推断总体，需要对样本值进行“加工”，这就要构造一些样本的函数(统计量)，它把样本中所含的（某一方面）的信息集中起来.

◆ 定义 设$X_1,X_2,\cdots,X_n$为总体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，不含未知参数的连续函数，则称$T(X_1,X_2,\cdots,X_n)$为一个统计量.

统计量

- 统计量是随机变量.
- 对给定的样本值$x_1,x_2,\cdots,x_n$可以计算出统计量$T(X_1,X_2,\cdots,X_n)$的值$T(x_1,x_2,\cdots,x_n)$.

例1 $X\sim N(\mu,\sigma^2), X_1,\cdots,X_n$ 为来自总体X的样本，μ未知，σ^2已知，判断下列函数那些是统计量.

$$\frac{1}{n}\sum_{i=1}^{n}X_i \qquad \cancel{\frac{1}{n}\sum_{i=1}^{n}X_i-\mu} \qquad \max(X_1,\cdots,X_n)$$

常用统计量

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

用途：当总体均值$E(X)=\mu$未知时，可用样本均值$\bar{X}$去估计总体均值μ.

2. 样本方差

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

常用统计量

- 样本标准差

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

用途：当总体方差$D(X)=\sigma^2$未知时，可用样本方差S^2去估计总体方差σ^2.

用样本标准差S去估计总体标准差σ.

常用统计量

3. 样本k阶原点矩

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

$$\bar{X}=A_1=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

用途：当总体k阶原点矩$E(X^k)=\alpha_k$未知时，可用样本k阶原点矩A_k去估计总体k阶原点矩α_k.

常用统计量

4. 样本k 阶中心原点矩

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

用途：当总体k 阶中心矩$E[X-E(X)]^k=\beta_k$未知时，可用样本k 阶中心矩B_k去估计总体k阶中心矩β_k.

常用统计量

- 样本的2阶中心矩用S^{*2}表示

$$S^{*2}=B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2.$$

用S^{*2}可以估计总体方差$D(X)$.

样本方差S^2与样本2阶中心矩S^{*2}间的关系为

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

常用统计量

5. 顺序统计量

设$X_1, X_2, \cdots X_n$为总体X的一个样本，$x_1, x_2, \cdots x_n$是样本值，将它们按大小次序排列，得

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}.$$

称$X_{(i)}$为第i个顺序统计量，如果不论样本$X_1, X_2, \cdots X_n$取哪组观测值$x_1, x_2, \cdots, x_n$，$X_{(i)}$总是取$x_{(i)}$为观测值$(i = 1, 2, \cdots, n)$.

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)},$$

常用统计量

$$X_{(1)}=\min(X_1,\dots,X_n)\text{为最小顺序统计量,}$$

$$X_{(n)}=\max(X_1,\dots,X_n)\text{为最大顺序统计量.}$$

6. 样本中位数

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

7. 样本极差

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

抽样分布

- 当用统计量推断总体时，必须知道统计量的分布，统计量的分布属于样本函数的分布，人们把**样本函数的分布统称为抽样分布**.

下面分别给出**单个正态总体**和**两个正态总体**统计量的分布.

单个正态总体统计量的分布

定理1（样本均值的分布）

设$X_1, X_2, \cdots, X_n$为总体$N(\mu, \sigma^2)$的一个样本，

样本均值$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$，则 $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$.

证 由于$\bar{X}$是n个独立正态变量$X_1, \cdots, X_n$的线性组合，因此$\bar{X} \sim N(E(\bar{X}), D(\bar{X}))$.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\mu = \mu,$$

单个正态总体统计量的分布

$$D(\bar{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

故 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

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

$$\frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2(1).$$

单个正态总体统计量的分布

定理2(样本方差的分布)

设$X_1, X_2, \cdots, X_n$为总体$N(\mu, \sigma^2)$的一个样本，

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

(1)样本方差S^2与样本均值$\bar{X}$相互独立；

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

单个正态总体统计量的分布

推论

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

$\sum_{i=1}^{n}(X_i-\bar{X})^2$有$n-1$个自由度, 因为有一个约束

条件$\sum_{i=1}^{n}(X_i-\bar{X})=0.$

单个正态总体统计量的分布

定理3 设$X_1, X_2, \cdots, X_n$为总体$N(\mu, \sigma^2)$的一个样本，$\bar{X}$和S^2分别为样本均值和样本方差，

则 $$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1).$$

证 $$\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

$\bar{X}$和S^2独立, 则$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} \sim t(n-1).$$

单个正态总体统计量的分布

例2 设$X_1, X_2, \cdots, X_n$为总体X的一个容量为n的样本，$E(X)=\mu, D(X)=\sigma^2>0$,样本均值

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

求（1）$E(\bar{X}), D(\bar{X}), E(S^2)$

（2）若$X \sim N(\mu,\sigma^2)$,则$D(S^2)$.

解

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\mu = \mu,$$

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

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

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

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

$$=\frac{1}{n-1}\left\{n[D(X)+(E(X))^2]-n[D(\bar{X})+(E(\overline{X}))^2]\right\}$$

$$=\frac{1}{n-1}\left\{n[\sigma^2+\mu^2]-n[\frac{\sigma^2}{n}+\mu^2]\right\}=\sigma^2.$$

结论：$E(\bar{X})=E(X), D(\bar{X})=\dfrac{D(X)}{n},$

$E(S^2)=D(X).$

(2)

总体$X \sim N(\mu,\sigma^2)$,有$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$$D\left[\frac{(n-1)S^2}{\sigma^2}\right]=2(n-1).$$

$$\frac{(n-1)^2 D(S^2)}{\sigma^4}=2(n-1).$$

$$D(S^2)=\frac{2\sigma^4}{n-1}.$$

两个正态总体统计量的分布

定理4 设$X_1,X_2,\cdots,X_{n_1}$和$Y_1,Y_2,\cdots,Y_{n_2}$分别是来自总体$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$的两个样本，它们**相互独立**，样本均值分别为$\bar{X}$和$\bar{Y}$,样本方差分别为S_1^2和S_1^2，则

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

两个正态总体统计量的分布

（2）$\sigma_1^2=\sigma_2^2=\sigma^2$时，

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

其中

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

两个正态总体统计量的分布

证(1)

$$\left.\begin{array}{l} \dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1), \\[2ex] \dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1), \end{array}\right\} \text{独立}$$

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

两个正态总体统计量的分布

证（2）$\sigma_1^2=\sigma_2^2=\sigma^2$时，

$$\bar{X}\sim N\left(\mu_1,\frac{\sigma^2}{n_1}\right),\bar{Y}\sim N\left(\mu_2,\frac{\sigma^2}{n_2}\right),$$ 它两独立，

$$\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right),$$

$$\bar{X}-\bar{Y}-(\mu_1-\mu_2)\Big/\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sim N(0,1).$$

两个正态总体统计量的分布

$$\frac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2(n_1-1),\quad \frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_2-1),$$

它两独立，由 χ^2 分布的可加性

$$\left.\begin{array}{l}\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_1+n_2-2).\\[2em] \bar{X}-\bar{Y}-(\mu_1-\mu_2)\Big/\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}\sim N(0,1).\end{array}\right\}\text{独立}$$

两个正态总体统计量的分布

由t分布定义 $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

$$=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\Bigg/\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2(n_1+n_2-2)}}$$

$$\sim t(n_1+n_2-2).$$

谢　谢！