第二节　统计量

一、统计量的概念

二、统计量的性质

6.2.1 统计量的概念

1. 定义

设 $(X_1, X_2, \dots, X_n)$ 是取自总体X的一个简单随机样本，若样本函数$T = T(X_1, X_2, \dots, X_n)$中不含任何未知参数，则称$T$为统计量。统计量的分布称为抽样分布。若$(x_1,\cdots,x_n)$为样本的观测值，则称T$(x_1,\cdots,x_n)$为统计量的观测值或统计值。

例如, 设总体 $X \sim N(\mu, \sigma 2)$, 参数未知,样本为 X_1 , X_2 . $T_1 = X_1 e^{X2}$ 是统计量,$T_2 = E（X_1 + X_2）$不是统计量.

Statistic

Q :If X_1, X_2, X_3 are the samples comes from Normal population $N(\mu, \sigma^2)$, μ is known, σ^2 is unknown. Which are statistics in the following expression?

$$T_1 = X_1, \qquad T_4 = \max(X_1, X_2, X_3),$$

$$T_2 = X_1 + X_2 e^{X_3}, \qquad T_5 = X_1 + X_2 - 2\mu,$$

$$T_3 = \frac{1}{3}(X_1 + X_2 + X_3), \qquad T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2).$$

6.2.1 统计量的概念

2. 常用统计量:

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

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

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

6.2.1 统计量的概念

样本k阶原点矩： $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

样本k阶中心矩： $B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k$

这七种样本数字特征统称为样本矩。

➢**Theorem 5.1** Suppose the *k*th sample moment of X exists, then $if\ n \to \infty,$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k),\ k = 1, 2, \cdots.$$

The property of sample mean and variance

Theorem 5.2: Suppose population X satisfied

$$E(X) = \mu, Var(X) = \sigma^2, \quad \text{then :}$$

(1) $E(\overline{X}) = \mu;$ (2) $Var(\overline{X}) = \frac{1}{n}\sigma^2;$

(3) $E(S^2) = \sigma^2;$

The property of sample mean and variance

Prove: $(1)\ E(\overline{X}) = \mu;$ $(2)\ Var(\overline{X}) = \frac{1}{n}\sigma^2;$

$(3)\ E(S^2) = \sigma^2;$

$$E(X) = E(X_i) = \mu, \quad Var(X) = Var(X_i) = \sigma^2$$

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

$$(2)\ \mathrm{Var}(\overline{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}\sigma^2$$

$$(3)\ E(S^2)=E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2]=\frac{1}{n-1}E(\sum_{i=1}^{n}X_i^2-n\overline{X}^2)$$

$$=\frac{1}{n-1}[\sum_{i=1}^{n}E(X_i^2)-nE(\overline{X}^2)]=\frac{1}{n-1}[nE(X_1^2)-nE(\overline{X}^2)]$$

$$=\frac{n}{n-1}[Var(X_1)+(EX_1)^2-(Var(\overline{X})+(E\overline{X})^2)]$$

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

$$=\sigma^2$$

The property of sample mean

Theorem 5.3: Suppose sample $X_1, \ldots, X_n$ are i.i.d. with the mean μ and variance σ^2, t>0 is a constant, then by Chebyshev inequality, the sample mean $\bar{X}$ satisfied:

$$P(|\bar{X}-\mu| \geq t) \leq \frac{\sigma^2}{nt^2}$$

例：确定所需的观察次数

假设n个随机样本取自平均值未知μ但方差已知σ^2=4的分布总体. 若至少有0. 99的概率$|\bar{X}-\mu|$ 取值小于1，确定样本量必须有多大。

解：对于大小为n的样本，通过切比雪夫不等式可知，

$$P(|\bar{X}-\mu| \geq 1) \leq \frac{\sigma^2}{n} = \frac{4}{n}.$$

$$P(|\bar{X}-\mu| < 1) = 1 - P(|\bar{X}-\mu| \geq 1) \geq 1 - \frac{4}{n} \geq 0.99$$

$$\Longrightarrow \quad \frac{4}{n} < 0.01 \Longrightarrow \quad n \geq 400.$$

3. 顺序统计量定义

设（$X_1,X_2,\ldots,X_n$）为总体X的一个样本，（$x_1,x_2,\ldots,x_n$）为样本的观测值，将它们按大小次序排列，得到$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$，如果不论样本观测值（$x_1,x_2,\ldots,x_n$）取何值，随机变量$X_{(i)}$总是取其中的$x_{(i)}$为观察值，称$X_{(i)}$为样本（$X_1,X_2,\ldots,X_n$）第$i$个**顺序统计量**，（$i=1,2,\ldots,n$）.

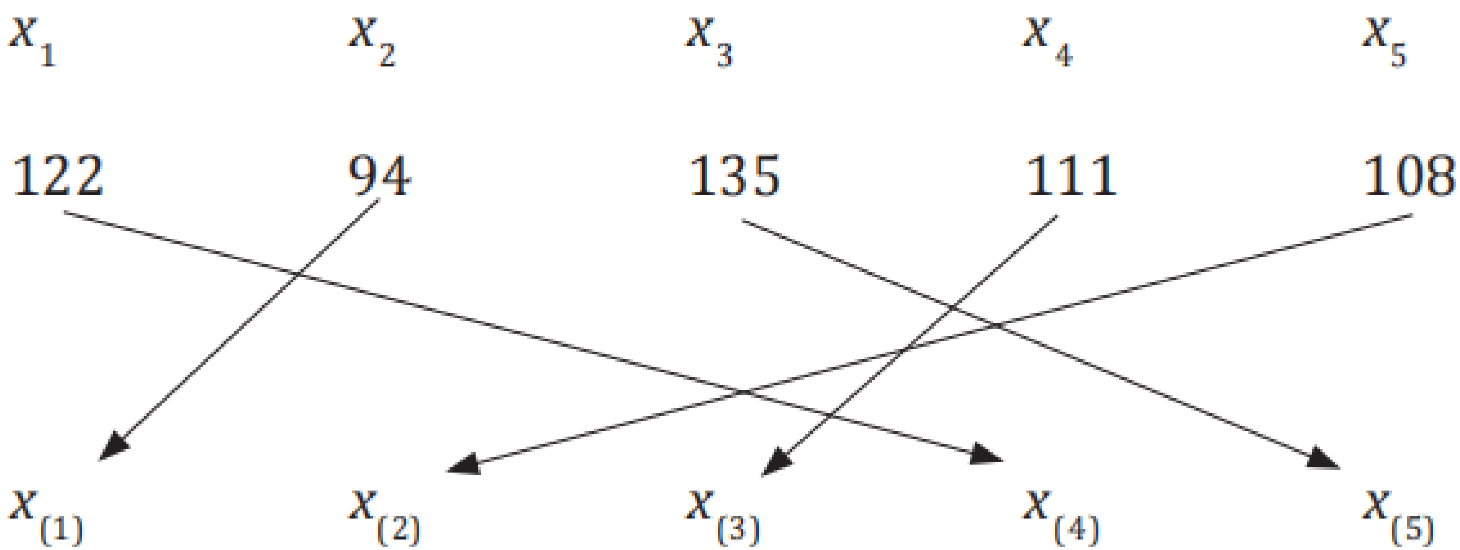
3. 顺序统计量定义

显然，$X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$，$X_{(1)}$和$X_{(n)}$常常分别称为**最小和最大顺序统计量**. 并称$X_{(n)}-X_{(1)}$为**样本极差**，其反映了总体分布的分散顺序。

➢Order statistic:

Rearrange the elements $\boldsymbol{x_1, x_2, \ldots, x_n}$ of the random sample $\boldsymbol{X_1, X_2, \ldots, X_n}$ in an increasing order $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ to yield r.v. $\boldsymbol{X_{(1)}, \ldots, X_{(n)}}$.

Example:



$\boldsymbol{X_{(1)}, \ldots, X_{(n)}}$ are order statistic.

So we know

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n)$$

$$X_{(n)} = \max(X_1, X_2, \cdots, X_n)$$

- $\boldsymbol{X}_{(\mathbf{1})}, \ldots, \boldsymbol{X}_{(\boldsymbol{n})}$ are also random variables and they are often not independent .

If $F(x)$ and $p(x)$ are the cdf and pdf of population X,

$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is the order statistic, then

(1) The pdf of $X_{(1)}$ is

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x)$$

(2) The pdf of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$$

Example: Suppose population X~U[0, θ], $(X_1, \ldots X_n)$ are samples. What is the pdf of $X_{(1)}$, $X_{(n)}$?

Solution: The pdf and cdf of X are

$$f(x)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta\\ 0, & else.\end{cases} \qquad F(x)=\begin{cases}0, & x<0\\ \dfrac{x}{\theta}, & 0\le x\le\theta\\ 1, & x>\theta\end{cases}$$

We get $X_{(1)}$'s pdf is

$$f_{X_{(1)}}(x)=\begin{cases}\dfrac{n}{\theta}(1-\dfrac{x}{\theta})^{n-1}, & 0\le x\le\theta\\ 0, & else\end{cases}$$

and $X_{(n)}$'s pdf is

$$f_{X_{(n)}}(x)=\begin{cases}\dfrac{n}{\theta^n}x^{n-1}, & 0\le x\le\theta\\ 0, & else\end{cases}$$

6.2.2 统计量的性质

例6.2.2 设总体的二阶矩存在，（$X_1,X_2,\ldots,X_n$）是样本。证明：

$X_i-\overline{X}$与$X_j-\overline{X}$的相关系数为$-\frac{1}{n-1}$,其中$i\neq j$.

证明 设总体X的方差为$D(X)=\sigma^2$.由定义，$X_i-\overline{X}$与$X_j-\overline{X}$的相关系数为

$$\rho=\rho\left(X_i-\overline{X},X_j-\overline{X}\right)=\frac{Cov(X_i-\overline{X},X_j-\overline{X})}{\sqrt{D(X_i-\overline{X})}\sqrt{D(X_j-\overline{X})}}$$

$$Cov\left(X_i-\overline{X},X_j-\overline{X}\right)=Cov\left(X_i,X_j\right)-Cov(X_i,\overline{X})-Cov\left(X_j,\overline{X}\right)+Cov(\overline{X},\overline{X})$$

6.2.2 统计量的性质

$$Cov\left(X_i-\bar{X},X_j-\bar{X}\right)=Cov\left(X_i,X_j\right)-Cov(X_i,\bar{X})-Cov\left(X_j,\bar{X}\ \right)+Cov(\bar{X},\bar{X}\ )$$

因为$X_1,X_2,\ldots,X_n$的独立性，则有

$$Cov\left(X_i,X_j\right)=0,\qquad Cov(\bar{X},\bar{X})=D(\bar{X})=\frac{\sigma^2}{n}$$

$$Cov(X_i,\bar{X})=Cov\left(X_j,\bar{X}\ \right)=\mathrm{Cov}\left(X_i,\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{\sigma^2}{n}$$

$$Cov\left(X_i-\bar{X},X_j-\bar{X}\right)=Cov\left(X_i,X_j\right)-Cov(X_i,\bar{X})-Cov\left(X_j,\bar{X}\ \right)+Cov(\bar{X},\bar{X}\ )$$

$$=-\frac{\sigma^2}{n}$$

6.2.2 统计量的性质

$$\sqrt{D(X_i-\overline{X})}=\sqrt{D(X_j-\overline{X})}=\sqrt{D(X_1-\overline{X})}$$

$$=\sqrt{D\left(\frac{(n-1)X_1-X_2-\cdots-X_n}{n}\right)}$$

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

所以 $$\rho=\rho\left(X_i-\overline{X},X_j-\overline{X}\right)=\frac{Cov(X_i-\overline{X},X_j-\overline{X})}{\sqrt{D(X_i-\overline{X})}\sqrt{D(X_j-\overline{X})}}$$

$$=-\frac{1}{n-1}$$

小结

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

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

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

样本k阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$　样本k阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$