第8讲 统计量及其分布

知识结构

- **统计量**
  - 样本均值 — $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$
  - 样本方差 — $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)$
  - 样本标准差 — $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$
  - 样本 k 阶原点矩 — $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\ (k=1,2,\cdots)$
  - 样本 k 阶中心矩 — $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k\ (k=2,3,\cdots)$
  - 顺序统计量 $\begin{cases}X_{(1)}=\min\{X_1,X_2,\cdots,X_n\}\\X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}\end{cases}$
  - 经验分布函数(仅数学三) — $F_n(x)=\begin{cases}0, & x<x_{(1)}\\ \frac{k}{n}, & x_{(k)}\leqslant x<x_{(k+1)}\ (k=1,2,\cdots,n-1)\\ 1, & x\geqslant x_{(n)}\end{cases}$
- **统计量的四大分布**
  - 正态分布 — $X\sim N(0,1),P\{X>\mu_\alpha\}=\alpha(0<\alpha<1)$
  - χ^2 分布 $\begin{cases}X_1,X_2,\cdots,X_n\text{ 独立同服从 }N(0,1)\text{,则 }\boxed{X=\sum_{i=1}^{n}X_i^2}\sim\chi^2(n)\\EX=n,DX=2n\end{cases}$
  - t 分布 $\begin{cases}X\sim N(0,1),Y\sim\chi^2(n),X,Y\text{ 独立},t=\frac{X}{\sqrt{Y/n}}\sim t(n)\\t_{1-\alpha}(n)=-t_\alpha(n)\end{cases}$
  - F 分布 $\begin{cases}X\sim\chi^2(n_1),Y\sim\chi^2(n_2),X,Y\text{ 独立},F=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)\\\frac{1}{F}\sim F(n_2,n_1),F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}\end{cases}$
- **正态总体下的常用结论**
  - ① $\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$,即 $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,1)$
  - ② $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$
  - ③ $\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\sim\chi^2(n-1)$
  - ④ $\frac{n(\overline{X}-\mu)^2}{S^2}\sim F(1,n-1)$

$\frac{X}{\sqrt{\chi^2(n)/n}}\sim t(n)$

称研究对象全体的某数量指标为总体 X，n 个相互独立且与总体 X 具有相同概率分布的随机变量 X_1，$X_2,\cdots,X_n$ 所组成的整体 $(X_1,X_2,\cdots,X_n)$ 称为来自总体 X 的容量为 n 的一个简单随机样本，简称样本．一次抽样结果的 n 个具体数值 $(x_1,x_2,\cdots,x_n)$ 称为样本 $X_1,X_2,\cdots,X_n$ 的一个观测值(或样本值)．

当 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个样本时，$g(x_1,x_2,\cdots,x_n)$ 为 n 元函数，如果 g 中不含任何未知参数，则称 $g(X_1,X_2,\cdots,X_n)$ 为样本 $X_1,X_2,\cdots,X_n$ 的一个**统计量**．统计量就是由统计数据计算得来的量．统计量是随机变量的函数，也是随机变量．

一 统计量

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的简单随机样本，则相应的统计量定义如下．

① **样本均值** $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$．

② **样本方差** $S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2\right)$；

样本标准差 $S=\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$．

③ **样本 k 阶原点矩** $A_k=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k\,(k=1,2,\cdots)$．

④ **样本 k 阶中心矩** $B_k=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k\,(k=2,3,\cdots)$．

⑤ **顺序统计量** 将样本 $X_1,X_2,\cdots,X_n$ 的 n 个观测量按其取值从小到大的顺序排列，得

$$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}.$$

随机变量 $X_{(k)}\,(k=1,2,\cdots,n)$ 称作第 k 顺序统计量，其中 $X_{(1)}$ 是最小观测量，而 $X_{(n)}$ 是最大观测量：

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

【注】数学三有一个考点，数学一不考，叫经验分布函数，即 $(x_1,x_2,\cdots,x_n)$ 为总体样本 $(X_1,X_2,\cdots,X_n)$ 的一个观测值，按大小顺序排列为 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$．对任意实数 x，称函数

$$F_n(x)=\frac{x_1,x_2,\cdots,x_n\text{ 中小于等于 }x\text{ 的样本值个数}}{n}$$

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

为样本 $(X_1,X_2,\cdots,X_n)$ 的**经验分布函数**．

事实上，$F_n(x)$ 就是事件 $\{X\leqslant x\}$ 在 n 次试验中出现的频率，而 $P\{X\leqslant x\}=F(x)$ 是事件 $\{X\leqslant x\}$ 出现的概率，由伯努利大数定律(即频率收敛于概率)可知，当 n 充分大时，$F_n(x)$ 可作为未知分布函数 $F(x)$ 的一个近似，n 越大，近似效果越好．

如设 $(2,1,5,2,1,3,1)$ 是来自总体 X 的简单随机样本值，求总体 X 的经验分布函数 $F_7(x)$．

解 将各观测值按从小到大的顺序排列，得 1,1,1,2,2,3,5，则经验分布函数为