6.样本及抽样分布

6.1总体和样本

1.总体

1.总体：具有一定共性的研究对象的全体集合

2.个体：：总体中的每一个元素

3.总体的分布：即随机变量的分布

2.样本

1.随机抽样：按照机会均等的原则，从总体X中抽取部分个体观察

2.样本：从总体X中抽取出的n个样本

  样本容量：样本所含个体的数目n

3.抽样方式：

  (1).不重复抽样

  (2).重复抽样

对于无限总体来说二者没有区别

4.简单随机抽样：重复抽样，每次抽样是独立同分布

6.2统计量

1.统计量的定义

定义：设$(X_1,X_2,X_3,\cdots,X_n)$是来自总体X的样本，设$g(x_1,x_2,x_2,\cdots,x_n)$为一个已知的$n$元实值函数，其中

不含有未知的参数，则$T=g(x_1,x_2,\cdots,x_n)$是一个随机变量，称为统计量

2.常用统计量

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

2.未修正的样本方差：$S_0^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

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

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

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

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

3.样本均值和样本方差的性质

1.设总体X的均值为$EX=\mu$，方差为$DX=\sigma^2$，样本$(X_1,X_2,\cdots,X_n)$来自总体X，则
(1) $E\overline{X}=\mu$:

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

(2)$D\overline{X}=\frac{1}{n}\sigma^2$:

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

(3)$ES^2=\sigma^2$:

6.3抽样分布

1. 重要分布

χ^2分布

1.设随机变量$X_1,X_2,X_3,\cdots,X_n$相互独立且均服从标准正态分布吗，则有

$$\chi^2=X_1^2+X_2^2+X_3^2+\cdots+X_n^2\quad \text{服从自由度为}n\text{的}\chi^2\text{分布，记为}\chi^2\sim\chi^2(n)$$

2.性质：

$$X\sim\chi^2(n)\text{，则}EX=n,DX=2n$$
$$\text{若}X\sim\chi^2(n)\text{，}Y\sim\chi^2(n)\quad\text{且}X,Y\text{相互独立，则}X+Y\sim\chi^2(n+m)$$

3.上α分位点

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

$$P\{\chi^2>\chi_\alpha^2(n)\}=\alpha$$

的点$\chi_\alpha^2(n)$为$\chi^2(n)$分布上的上α分位点

t分布

1.$X\sim N(0,1),\ Y\sim\chi^2(n),\ X,Y$独立，则称随机变量：

$$T=\frac{X}{\sqrt{Y/n}}$$

服从的分布为自由度为n的t分布，记作$T\sim t(n)$

可以证明，自由度无限增大时，t分布将趋于标准正态分布　$t\geq 30$ 时用正态分布近似

2.上α分位点

由概率密度的对称性知: $t_{1-\alpha}(n)=-t_\alpha$

F分布

1.$F(n_1,n_2)$

$X\sim\chi^2(n_1),Y\sim\chi^2(n_2)\quad X,Y\text{相互独立}\quad \frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$

因此有 若$F\sim F(n_1,n_2)$, 则$\frac{1}{F}\sim F(n_2,n_1)$

2.上α 分位数

$$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_1,n_2)}$$

2.正态总体下的抽样分布

总体为正态分布，抽取样本，求这些样本构造的统计量的分布

1.一个总体正态分布 $X\sim N(\mu,\sigma^2)\quad\{X_1,\cdots,X_n\}\text{为样本}$

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

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

(3). $\overline{X}\text{与}S^2\text{相互独立}$

(4). $\frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\mu)\sim\chi^2(n)$

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

证明:

$$\begin{aligned}&\overline{X}\sim N(\mu,\frac{\sigma^2}{n})\quad\text{标准化得到：}\frac{\overline{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)\\&\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)\\&\frac{\overline{X}-\mu}{\sigma}\sqrt{n}\cdot\sqrt{\frac{\sigma^2}{(n-1)S^2}}\cdot\sqrt{n-1}=\frac{\overline{X}-\mu}{S}\sqrt{n}\sim t(n-1)\end{aligned}$$

2.两个总体正态分布 $X\sim(\mu_1,\sigma_1^2)$, $Y\sim(\mu_2,\sigma_2^2)\quad\text{样本}\{X_1,\cdots,X_{n_1}\}$, $\{Y_1,\cdots,Y_{n_2}\}$

(1) $\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$

证明:

$$\begin{aligned}&\overline{X}\sim N(\mu_1,\frac{\sigma_1^2}{n_1}),\qquad\qquad \overline{Y}\sim N(\mu_2,\frac{\sigma_2^2}{n_2})\\&\overline{X}-\overline{Y}-(\mu_1-\mu_2)\sim N(0,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})\\&\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)\end{aligned}$$

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

证明:

$$
\begin{aligned}
Z_1 &= \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \mathcal{X}(n_1-1) \\
Z_2 &= \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \mathcal{X}(n_2-1) \\
\frac{Z_1/(n_1-1)}{Z_2/(n_2-1)} &= \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)
\end{aligned}
$$

（3）$\sigma_1^2=\sigma_1^2=\sigma^2$时，有$\frac{(\overline{X}-\overline{Y}-(\mu_1-\mu_2))}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-1)$

证明：

$$
\begin{aligned}
&S_w^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \\
&\overline{X}-\overline{Y}-(\mu_1-\mu_2) \sim N(0,(\frac{1}{n_1}+\frac{1}{n_2})\sigma^2) \\
&U = \frac{1}{\sigma\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\cdot(\overline{X}-\overline{Y}-(\mu_1-\mu_2)) \sim N(0,1) \\
&\frac{(n_1-1)S_1^2}{\sigma^2} \sim \mathcal{X}(n_1-1), \frac{(n_2-1)S_2^2}{\sigma^2} \sim \mathcal{X}(n_2-1) \\
&V = \frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \mathcal{X}(n_1+n_2-1) \\
&\frac{U}{\sqrt{V/(n_1+n_2-1)}} = \frac{(\overline{X}-\overline{Y}-(\mu_1-\mu_2))}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-1)
\end{aligned}
$$