统计学基本概念

13.3 常用统计量

统计量

设想你参加了一次考试，在知道自己得到了 78 分后，希望了解自己的成绩在班级上处于什么水平。你会怎样做？

你对自己未来工作收入的预期是什么？

定义：设 $X_1, X_2, \cdots, X_n$ 为取自某总体的样本，若样本函数 $T = T\left(X_1, X_2, \cdots, X_n\right)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布称为抽样分布。

****

强国知十三数：境内仓口之数，壮男壮女之数，老弱之数，官士之数，以言说取食者之数，利民之数，马牛刍藁之数。欲强国，不知国十三数，地虽利，民虽众，国愈弱至削。国无怨民曰强国。兴兵而伐，则武爵武任，必胜；按兵而农，粟爵粟任，则国富。兵起而胜敌，按兵而国富者，王。

（秦·商鞅《商君书》）

商鞅（前 390～前 338 年），卫国家，思想家，著名法家代表人物。应秦孝公求贤令入秦，说服秦孝公变法图强。孝公死后，受到贵族诬害以及秦惠文王的猜忌，车裂而死。其在秦执政二十余年，秦国大治，史称“商鞅变法”。

****

统计量是对样本的一种加工。常用的统计量有样本均值、样本方差等。

定义 设$X_1,X_2,\cdots,X_n$为取自某总体的样本，则$\bar{X}=\dfrac{X_1+X_2+\cdots+X_n}{n}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$称为样本均值。

定理 设$X_1,X_2,\cdots,X_n$是来自某个总体X的样本，$\bar{X}$为样本均值，

（1）若总体$X\sim N(\mu,\sigma^2)$，则$\bar{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$；

证明：　$X_1,X_2,\cdots,X_n$相互独立，$X_k\sim N(\mu,\sigma^2)$　$k=1,2,\cdots n$

$$E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n}=\frac{n\mu}{n}=\mu$$

$$Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{Var(X_1)+Var(X_2)+\cdots+Var(X_n)}{n^2}=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

（2）若总体分布不是正态分布，已知 $E(X)=\mu$，$D(X)=\sigma^2$，则n较大时，$\bar{X}$的渐近分布为$N\left(\mu,\dfrac{\sigma^2}{n}\right)$，常记为$\bar{X}\dot{\sim} N\left(\mu,\dfrac{\sigma^2}{n}\right)$。

****

定义 设$X_1,X_2,\cdots,X_n$是来自某个总体X的样本，$\bar{X}$为样本均值，则

$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$称为样本方差。

定理　设总体X具有二阶中心矩，$E(X)=\mu$，$Var(X)=\sigma^2<+\infty$，$X_1,X_2,\cdots,X_n$为来自该总体的样本，$\bar{X}$和S^2分别是样本均值和样本方差，则$E\left(S^2\right)=\sigma^2$。

样本方差是总体方差的无偏估计，样本均值是总体期望的无偏估计。

****

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

证明： $E\left(\bar{X}\right)=\mu$， $Var\left(\bar{X}\right)=\dfrac{\sigma^2}{n}$，

$$E\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=E\sum_{i=1}^{n}\left[\left(X_i-E\left(X_i\right)\right)-\left(\bar{X}-E\left(\bar{X}\right)\right)\right]^2$$

$$=\sum_{i=1}^{n}E\left(X_i-E\left(X_i\right)\right)^2+\sum_{i=1}^{n}E\left(\bar{X}-E\left(\bar{X}\right)\right)^2-2\cdot E\sum_{i=1}^{n}\left[\left(X_i-E\left(X_i\right)\right)\cdot\left(\bar{X}-E\left(\bar{X}\right)\right)\right]$$

$$=\sum_{i=1}^{n}Var\left(X_i\right)+\sum_{i=1}^{n}Var\left(\bar{X}\right)-2\cdot E\left[\left(\sum_{i=1}^{n}X_i-nE\left(X_i\right)\right)\cdot\left(\bar{X}-E\left(\bar{X}\right)\right)\right]$$

$$=n\sigma^2+n\cdot Var\left(\bar{X}\right)-2\cdot E\left[\left(n\bar{X}-nE\left(\bar{X}\right)\right)\cdot\left(\bar{X}-E\left(\bar{X}\right)\right)\right]$$

$$=n\sigma^2-n\cdot Var\left(\bar{X}\right)=\left(n-1\right)\sigma^2.$$

****

其他常用的统计量

设 $X_1,X_2,\cdots,X_n$ 是来自某个总体 X 的样本

样本 k 阶原点矩 $a_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$

样本 k 阶中心矩 $b_k=\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^k$，

其中 $X_1,X_2,\cdots,X_n$ 为来自总体的样本，$\bar{X}$ 为样本均值。

****