第六章 统计量及其抽样分布

6.1 统计量

统计量的定义：如果 $X_1, X_2, ..., X_n$ 是总体 X 中抽取的容量为 n 的一个样本，如果不依赖于其他未知参数构造一个函数 $T(X_1, X_2, ..., X_n)$，则称这个函数为一个（样本）统计量。代入具体值 $(x_1, x_2, ..., x_n)$ 可以获得具体的统计量。统计量就是只由统计获得的，你不知道其他的关于总体的量（例如平均、方差等）但是用样本平均之类的是可以的。

> 常用统计量：
>
> 样本平均 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
>
> 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
>
> 样本变异系数 $V = S/\overline{X}$
>
> 样本 k 阶矩 $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
>
> 样本 k 阶中心矩 $v_k = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
>
> 样本偏度反映随机变量的密度函数曲线在众数两边的偏斜性。正态分布的偏度为 0。
>
> $$\alpha_3 = \sqrt{n-1}\frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{\sum_{i=1}^{n}(X_i - \overline{X})^{3/2}} \tag{1}$$
>
> 样本峰度反映随机变量的密度函数曲线在众数附近的"峰"的尖峭程度。正态分布的峰度为 0。
>
> $$\alpha_4 = \frac{(n-1)\sum_{i=1}^{n}(X - \overline{X})^4}{[\sum_{i=1}^{n}(X_i - \overline{X})^2]^2 - 3} \tag{2}$$

6.2 由正态分布导出的一些重要分布

6.2.1 抽样分布

在 X 的总体分布已知时，若对于任一自然数 n，都能导出统计量 $T=T(X_1,X_2,\ldots,X_n)$ 的分布的数学表达式，这种分布就是精确的抽样分布。在正态总体的条件下，主要由 χ^2，t 和 F 三大分布。

6.2.2 χ^2 分布（卡方分布）

设$X_1,X_2,\ldots,X_n$ 相互独立，且 $X_i(i=1,2,\ldots,n)$ 服从标准正态分布$\mathcal{N}(0,1)$，则平方和 $\sum_{i=1}^n X_i^2$ 服从自由度为 n 的 χ^2 分布。

- χ^2 分布的均值为 $E(\chi^2)=n$，方差为 $D(\chi^2)=2n$。
- χ^2 分布具有可加性，$\chi_1^2\sim\chi^2(n_1),\chi_2^2\sim\chi^2(n_2)$，则 $\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$
- χ^2 分布的 p 分位数 q（$P(X\leq q)=p$）由查表知道
  - 当自由度很大时，$\sqrt{2\chi^2}\sim\mathcal{N}(\sqrt{2n-1},1)$;
  - 当自由度$n>45$ $\chi_p^2(n)\approx\frac{1}{2}\left(\mu_p+\sqrt{2n-1}\right)^2$，$\mu_p$ 是正态 p 分位数。

6.2.3 t 分布（学生分布）

t 分布对统计学小样本理论及其应用有贡献。设随机变量 $X\sim\mathcal{N}(0,1)$，$Y\sim\chi^2(n)$，且X,Y 相互独立，则

$$t=\frac{X}{\sqrt{Y/n}} \tag{3}$$

t 分布的目睹函数尾部比$\mathcal{N}(0,1)$ 粗，$t(n)$ 的方差更大。

- $n=1$ 的 t 分布是柯西分布，随着 n 增大越来越接近$\mathcal{N}(0,1)$;
- $n\geq 2$ 的 t 分布数学期望 $E(X)=0$;
- $n\geq 3$ 的 t 分布方差 $D(t)=\frac{n}{n-2}$。

> t 分布有关的抽样分布，见书本P120
>
> 设 $X_1,X_2,\ldots,X_n$ 是来自正态分布 $\mathcal{N}(\mu,\sigma)$ 的一个样本，$\overline{X}=\sum_{i=1}^n X_i$，$S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})^2$,则
>
> $$\frac{\sqrt{n}(\overline{X}-\mu)}{S}\sim t(n-1) \tag{4}$$

6.2.4 F 分布

F 分布应用于方差分析、回归方程的显著性检测。设随机变量 Y,Z 相互独立，且 Y,Z 分别服从自由度为 m 和 n 的 χ^2 分布，随机变量 $X\sim F(m,n)$，m 为第一自由度，n 为第二自由度。

$$
\begin{aligned}
X &= \frac{Y}{m} \cdot \frac{n}{Z} = \frac{nY}{mZ} \\
E(X) &= \frac{n}{n-2} && ,n \geq 2 \\
D(X) &= \frac{2n^2(m+n-2)}{m(n-2)(n-4)} && ,n \geq 4 \\
F_p(v_1, v_2) &= \frac{1}{F_{1-p}(v_2, v_1)} && ,p\text{分位数（还是一个X）}
\end{aligned} \tag{5}
$$

6.3 样本均值的分布与中心极限定理

$$
\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{6}
$$

当总体分布是正态分布$\mathcal{N}(\mu, \sigma)$ 的时候，$\overline{X}$ 的抽样仍是正态分布，$\overline{X}$ 的期望 μ，σ^2/n。

$$
\begin{aligned}
E(\overline{X}) &= E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i)\right) = \mu \\
D(\overline{X}) &= D\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} D(X_i)\right) = \frac{\sigma^2}{n}
\end{aligned} \tag{7}
$$

中心极限定理 设均值为 μ，方差位 σ^2 （有限）的任意一个中体中抽取样本为 n 的样本，当 n 充分大的时候，样本均值为 μ，方差为 σ^2/n 的正态分布。（一般 $n \geq 30$ 就是大样本）。