概率论与数理统计

统计量与经验分布函数

主讲人：郑旭玲

信息科学与技术学院

01

统 计 量

一、统计量

$$\text{样本} \xrightarrow{\text{统计量}} \text{统计推断}$$

不含任何未知参数的样本的函数称为统计量。

它是完全由样本决定的量。

一、统计量

定义

设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的函数，若g中不含未知参数，则称$g(X_1, X_2, \cdots, X_n)$是一个**统计量**。

设$X_1, X_2, \cdots X_n$是来自总体X的一个样本，$x_1, x_2, \cdots x_n$是一个样本的观察值，则$g(x_1, x_2, \cdots x_n)$是**统计量$g(X_1, X_2, \cdots X_n)$的观察值**。

一、统计量

例 设$X_1, \ldots, X_n$为来自总体 $X \sim N(\mu, \sigma^2)$ 的一个样本，其中μ 未知，σ^2已知，问下列随机变量中哪些是统计量？

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

[4] $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ [5] $X_1^2+X_2^2+\sigma^2$ [6] $2\mu X_1X_2\ldots X_n$

一、统计量

【几个常见统计量】

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

它反映了总体均值的信息

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

它反映了总体方差的信息

一、统计量

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

$$=\frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)$$

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

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

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

一、统计量

【几个常见统计量】

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

它反映了总体均值的信息

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

它反映了总体方差的信息

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

一、统计量

【几个常见统计量】

样本 k 阶原点矩 $$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\quad k = 1, 2, \ldots$$

它反映了总体k
阶矩的信息

其中，$A_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

样本 k 阶中心矩 $$B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k,\quad k = 2, 3, \ldots$$

它反映了总体k 阶
中心矩的信息

【辨析】 $B_2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$ $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$

一、统计量

【样本方差S^2的期望】

设总体X（不管服从什么分布，只要均值和方差存在）的均值为μ，方差为σ^2，$X_1, X_2, \ldots, X_n$是来自X的一个样本，$\bar{X}, S^2$分别是样本均值和样本方差，则有

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

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

一、统计量

【样本方差S^2的期望】

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

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

$$\because E(X_i^2)=D(X_i)+(EX_i)^2=\sigma^2+\mu^2$$

$$E[(\bar{X})^2]=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2$$

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

【辨析】 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$$E(B_2)=\frac{n-1}{n}\sigma^2$$

一、统计量

【统计量的观察值】 将样本观察值代入，便可求得统计量的观察值。

样本平均值 $\overline{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$

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

$$\alpha_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\quad k=1,2,\cdots$$

样本k阶中心矩

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\ k=2,3,\cdots$$

02

经验分布函数

二、经验分布函数

定义

设$X_1, X_2, \cdots, X_n$是总体F的一个样本，用$s(x), -\infty < x < \infty$表示$X_1, X_2, \cdots, X_n$中不大于$x$的随机变量的个数。

经验分布函数$F_n(x)$为

$$F_n(x) = \frac{1}{n} s(x), -\infty < x < \infty$$

对于一个样本，经验分布函数 $F_n(X)$的观察值（仍以 $F_n(X)$表示）是很容易得到的。

二、经验分布函数

例 下面给出了84个伊特拉斯坎（Etruscan）人男子的头颅的最大宽度（mm）。

141	148	132	138	154	142	150	146	155	158	150	140
147	148	144	150	149	145	149	158	143	141	144	144
126	140	144	142	141	140	145	135	147	146	141	136
140	146	142	137	148	154	137	139	143	140	131	143
141	149	148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138	142	149
142	137	134	144	146	147	140	142	140	137	152	145

二、经验分布函数

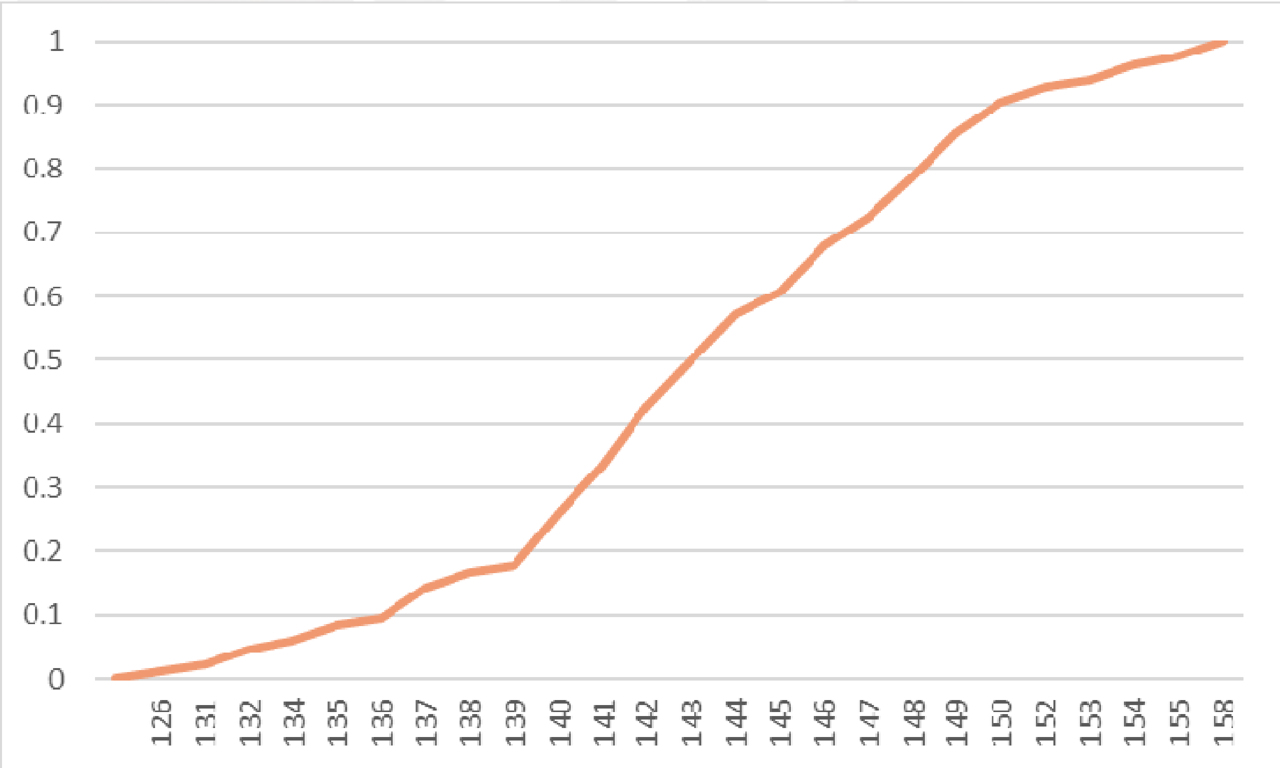
计算经验分布函数观察值的步骤：

1. 对样本数据从小到大进行排序，合并相同数据，并统计频数；
2. 用频数除以总数计算频率值；
3. 计算累积频率。

二、经验分布函数

数据	频数	频率	累积频率	区间	数据	频数	频率	累积频率	区间
			0	x<128	143	6	0.071429	0.5	143≤x<144
126	1	0.011905	0.011905	126≤x<131	144	6	0.071429	0.571429	144≤x<145
131	1	0.011905	0.02381	131≤x<132	145	3	0.035714	0.607143	145≤x<146
132	2	0.02381	0.047619	132≤x<134	146	6	0.071429	0.678571	146≤x<147
134	1	0.011905	0.059524	134≤x<135	147	4	0.047619	0.72619	147≤x<148
135	2	0.02381	0.083333	135≤x<136	148	5	0.059524	0.785714	148≤x<149
136	1	0.011905	0.095238	136≤x<137	149	6	0.071429	0.857143	149≤x<150
137	4	0.047619	0.142857	137≤x<138	150	4	0.047619	0.904762	150≤x<152
138	2	0.02381	0.166667	138≤x<139	152	2	0.02381	0.928571	152≤x<153
139	1	0.011905	0.178571	139≤x<140	153	1	0.011905	0.940476	153≤x<154
140	7	0.083333	0.261905	140≤x<141	154	2	0.02381	0.964286	154≤x<155
141	6	0.071429	0.333333	141≤x<142	155	1	0.011905	0.97619	155≤x<158
142	8	0.095238	0.428571	142≤x<143	158	2	0.02381	1	158≤x

二、经验分布函数



经验分布函数图

二、经验分布函数

一般，设$x_1, x_2, \cdots, x_n$是总体的一个容量为 n 的样本值.

将它们按大小次序排列如下：$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

则经验分布函数$F_n(x)$的观察值为

$$F_n(x) = \begin{cases} 0, & \text{若 } x < x_{(1)} \\ \dfrac{k}{n}, & \text{若 } x_{(k)} \le x < x_{(k+1)}, (k = 1, 2, \cdots, n-1) \\ 1, & \text{若 } x \ge x_{(n)} \end{cases}$$

二、经验分布函数

【格里汶科（Ghivenko）定理】

对于任一实数 x, 当 $n\to\infty$ 时, $F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$, 即

$$P\left\{\lim_{n\to\infty}\sup_{-\infty<x<+\infty}\left|F_n(x)-F(x)\right|=0\right\}=1.$$

对于任一实数 x 当 n 充分大时, 经验分布函数的任一个观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别, 从而在实际上可当作 $F(x)$ 来使用.

小结

- 统计量
- 经验分布函数

谢谢大家