2

数理统计

2.1样本及抽样分布

♀ 随机样本

我们将试验的全部可能观察值称为总体，这些值不一定都不相同，数目上也不一定有限，每一个可能观察值称为个体，总体包含个体的个数称为总体的容量。容量有限的总体成为有限总体，容量无限的成为无限总体。

定义： 设X是具有分布函数F的随机变量，若 $X_1, X_2, ..., X_n$ 是具有统一分布函数F的相互独立的随机变量，则称 $X_1, X_2, ..., X_n$为分布函数F(或总体F、或总体X)得到的容量为n的简单随机样本，简称样本，它们的观察值 $x_1, x_2, ..., x_n$称为样本值，又称为X的n个独立观察值。

♀ 直方图和箱线图

■ 频率直方图

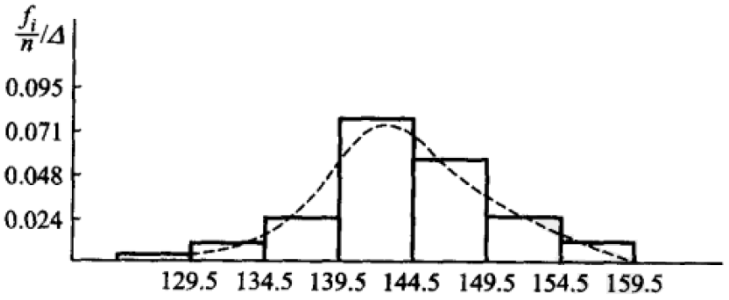
原始数据：

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						

分段分析：$\Delta = \frac{159.5 - 124.5}{7} = 5$

组　　限	频　数 f_i	频率 f_i/n	累积频率
124.5～129.5	1	0.011 9	0.011 9
129.5～134.5	4	0.047 6	0.059 5
134.5～139.5	10	0.119 1	0.178 6
139.5～144.5	33	0.392 9	0.571 5
144.5～149.5	24	0.285 7	0.857 2
149.5～154.5	9	0.107 1	0.952 4
154.5～159.5	3	0.035 7	1

频率直方图



■ 箱线图

定义：设有容量为n的样本观察值 $x_1, x_2, ..., x_n$ 样本p分位数 （$0 < p < 1$） 记为 x_p.

分位数的性质：

▲ 至少有n p个观察值小于或等于x_p;

▲ 至少有n (1 − p) 个观察值大于或等于x_p;

分位数的取值取决于n p的值： $x_p = \begin{cases} x_{([np]+1)} & np\text{不是整数} \\ \frac{1}{2}[x_{(np)} + x_{(np+1)}] & np\text{是整数} \end{cases}$

当p = 0.5时，0.5的分位数$x_{0.5}$也记作 Q_2或M 称为样本中位数；类似地 $x_{0.25}$称为第一四分位数记作 Q_1; $x_{0.75}$称第三四分位数记作Q_3

例 样本：

122	126	133	140	145	145	149	150	157
162	166	175	177	177	183	188	199	212

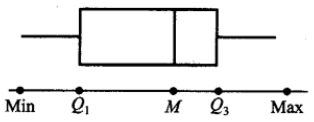
，求 $x_{0.2}, Q_1, Q_2$

1. np=18×0.2=3.6; [3.6]+1=4; $\Longrightarrow x_{0.2} = x_{(4)} = 140$

2. np = 18 × 0.25 = 4.5; [4.5] + 1 = 5; $\Longrightarrow x_{0.25} = x_{(5)} = 145$

3. np=18×0.5=9; $x_{0.5} = \frac{1}{2}(x_{(9)} + x_{(10)}) = \frac{1}{2}(157 + 162) = 159.5$

数据集的箱线图由箱子和直线组成图形基于五个内容： 最小值Min，第一四分位数 Q_1， 中位数M， 第三四分位数 Q_3 和最大数值Max



抽样分布

定义：设$X_1, X_2, ..., X_n$ 时来自总体X的一个样本， g($X_1, X_2, ..., X_n$)是$X_1, X_2, ..., X_n$ 的函数，若g中不含未知参数则称g($X_1, X_2, ..., X_n$)是一统计量。

几个常用统计量：

- 样本平均值： $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- 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： $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$
- 样本k阶(原点)距： $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1, 2, ...$
- 样本k阶中心距： $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k,\ k = 1, 2, ...$