

## 第六章 样本及抽样分布

1. 在总体  $N(52, 6.3^2)$  中随机抽取一容量为 36 的样本, 求样本均值  $\bar{X}$  落在 50.8 到 53.8 之间的概率.

解  $n=36, \bar{X} = \frac{1}{36} \sum_{i=1}^{36} X_i$ , 因总体  $X \sim N(52, 6.3^2)$ , 故  $E(\bar{X}) = 52, D(\bar{X}) = 6.3^2/36 = 1.05^2, \bar{X} \sim N(52, 1.05^2)$ . 从而

$$\begin{aligned} P\{50.8 < \bar{X} < 53.8\} &= P\left\{\frac{50.8 - 52}{1.05} < \frac{\bar{X} - 52}{1.05} < \frac{53.8 - 52}{1.05}\right\} \\ &= \Phi\left(\frac{53.8 - 52}{1.05}\right) - \Phi\left(\frac{50.8 - 52}{1.05}\right) \\ &= \Phi(1.71) - \Phi(-1.14) \\ &= \Phi(1.71) + \Phi(1.14) - 1 = 0.8293. \end{aligned}$$

2. 在总体  $N(12, 4)$  中随机抽一容量为 5 的样本  $X_1, X_2, X_3, X_4, X_5$ .

(1) 求样本均值与总体均值之差的绝对值大于 1 的概率.

(2) 求概率  $P\{\max\{X_1, X_2, X_3, X_4, X_5\} > 15\}, P\{\min\{X_1, X_2, X_3, X_4, X_5\} < 10\}$ .

解 (1)  $\bar{X} = \frac{1}{5} \sum_{i=1}^5 X_i$ , 因总体  $X \sim N(12, 4)$ , 故  $\bar{X} \sim N(12, 4/5)$ , 从而

$$\begin{aligned} P\{|\bar{X} - 12| > 1\} &= 1 - P\{|\bar{X} - 12| \leq 1\} \\ &= 1 - P\{-1 \leq \bar{X} - 12 \leq 1\} \\ &= 1 - P\left\{-\frac{1}{\sqrt{4/5}} \leq \frac{\bar{X} - 12}{\sqrt{4/5}} \leq \frac{1}{\sqrt{4/5}}\right\} \\ &= 1 - \left[\Phi\left(\frac{1}{\sqrt{4/5}}\right) - \Phi\left(\frac{-1}{\sqrt{4/5}}\right)\right] = 2 - 2\Phi(1.12) \\ &= 2(1 - 0.8686) = 0.2628. \end{aligned}$$

(2) 因  $X_i$  的分布函数为  $\Phi\left(\frac{x-12}{2}\right)$ , 故  $M = \max\{X_1, X_2, X_3, X_4, X_5\}$  的分布函数为

$$F_M(x) = \left[\Phi\left(\frac{x-12}{2}\right)\right]^5,$$

因而



$$\begin{aligned}
 & P\{\max\{X_1, X_2, X_3, X_4, X_5\} > 15\} \\
 &= P\{M > 15\} \\
 &= 1 - P\{M \leq 15\} = 1 - F_M(15) = 1 - \left[\Phi\left(\frac{15-12}{2}\right)\right]^5 \\
 &= 1 - 0.9332^5 = 0.2923.
 \end{aligned}$$

记  $N = \min\{X_1, X_2, X_3, X_4, X_5\}$ , 则  $N$  的分布函数为

$$F_N(x) = 1 - \left[1 - \Phi\left(\frac{x-12}{2}\right)\right]^5,$$

故

$$\begin{aligned}
 & P\{\min\{X_1, X_2, X_3, X_4, X_5\} < 10\} \\
 &= P\{N < 10\} \\
 &= 1 - \left[1 - \Phi\left(\frac{10-12}{2}\right)\right]^5 = 1 - [1 - \Phi(-1)]^5 \\
 &= 1 - [\Phi(1)]^5 = 1 - (0.8413)^5 = 0.5785.
 \end{aligned}$$

3. 求总体  $N(20, 3)$  的容量分别为 10, 15 的两独立样本均值差的绝对值大于 0.3 的概率.

解 将总体  $N(20, 3)$  的容量分别为 10, 15 的两独立样本的均值分别记作  $\bar{X}, \bar{Y}$ , 则  $\bar{X} \sim N(20, 3/10), \bar{Y} \sim N(20, 3/15)$ , 从而

$$\bar{X} - \bar{Y} \sim N(20 - 20, 3/10 + 3/15),$$

即

$$\bar{X} - \bar{Y} \sim N(0, 1/2),$$

故所求概率为

$$\begin{aligned}
 p &= P\{|\bar{X} - \bar{Y}| > 0.3\} = 1 - P\{|\bar{X} - \bar{Y}| \leq 0.3\} \\
 &= 1 - P\left\{\frac{-0.3}{\sqrt{1/2}} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{1/2}} \leq \frac{0.3}{\sqrt{1/2}}\right\} \\
 &= 2 - 2\Phi(0.42) = 2(1 - 0.6628) = 0.6744
 \end{aligned}$$

注: 注意  $D(\bar{X} - \bar{Y}) = D(\bar{X}) + D(\bar{Y})$ .

4. (1) 设样本  $X_1, X_2, \dots, X_6$  来自总体  $N(0, 1)$ ,  $Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2$ , 试确定常数  $C$  使  $CY$  服从  $\chi^2$  分布.

(2) 设样本  $X_1, X_2, \dots, X_5$  来自总体  $N(0, 1)$ ,  $Y = \frac{C(X_1 + X_2)}{(X_3^2 + X_4^2 + X_5^2)^{1/2}}$ , 试确定常数  $C$  使  $Y$  服从  $t$  分布.

(3) 已知  $X \sim t(n)$ , 求证  $X^2 \sim F(1, n)$ .

解 (1) 因  $X_1, X_2, \dots, X_6$  是总体  $N(0, 1)$  的样本, 故

$$X_1 + X_2 + X_3 \sim N(0, 3), \quad X_4 + X_5 + X_6 \sim N(0, 3),$$

且两者相互独立. 因此



$$\frac{X_1 + X_2 + X_3}{\sqrt{3}} \sim N(0, 1), \quad \frac{X_4 + X_5 + X_6}{\sqrt{3}} \sim N(0, 1),$$

且两者相互独立. 按  $\chi^2$  分布的定义

$$\frac{(X_1 + X_2 + X_3)^2}{3} + \frac{(X_4 + X_5 + X_6)^2}{3} \sim \chi^2(2),$$

即  $\frac{1}{3}Y \sim \chi^2(2)$ , 即知  $C = \frac{1}{3}$ .

(2) 因  $X_1, X_2, \dots, X_5$  是总体  $N(0, 1)$  的样本, 故  $X_1 + X_2 \sim N(0, 2)$ , 即有

$$\frac{X_1 + X_2}{\sqrt{2}} \sim N(0, 1).$$

而

$$X_3^2 + X_4^2 + X_5^2 \sim \chi^2(3).$$

且  $\frac{X_1 + X_2}{\sqrt{2}}$  与  $X_3^2 + X_4^2 + X_5^2$  相互独立, 于是

$$\frac{(X_1 + X_2)/\sqrt{2}}{\sqrt{(X_3^2 + X_4^2 + X_5^2)/3}} = \sqrt{\frac{3}{2}} \frac{X_1 + X_2}{(X_3^2 + X_4^2 + X_5^2)^{1/2}} \sim t(3),$$

因此所求的常数  $C = \sqrt{\frac{3}{2}}$ .

(3) 按定义  $X \sim t(n)$ , 故  $X$  可表示成

$$X = \frac{Z}{\sqrt{Y/n}},$$

其中,  $Y \sim \chi^2(n)$ ,  $Z \sim N(0, 1)$  且  $Z$  与  $Y$  相互独立, 从而

$$X^2 = \frac{Z^2}{Y/n}.$$

由于  $Z \sim N(0, 1)$ ,  $Z^2 \sim \chi^2(1)$ , 上式右端分子  $Z^2 \sim \chi^2(1)$ , 分母中  $Y \sim \chi^2(n)$ , 又由  $Z$  与  $Y$  相互独立, 知  $Z^2$  与  $Y$  相互独立. 按  $F$  分布的定义得

$$X^2 \sim F(1, n).$$

5. (1) 已知某种能力测试的得分服从正态分布  $N(\mu, \sigma^2)$ , 随机取 10 个人参与这一测试. 求他们得分的联合概率密度, 并求这 10 个人得分的平均值小于  $\mu$  的概率.

(2) 在(1)中设  $\mu = 62$ ,  $\sigma^2 = 25$ , 若得分超过 70 就能得奖, 求至少有一人得奖的概率.

解 (1) 10 个人的得分分别记为  $X_1, X_2, \dots, X_{10}$ . 它们的联合概率密度为

$$f(x_1, x_2, \dots, x_{10}) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}},$$

$$\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i \sim N\left(\mu, \frac{\sigma^2}{10}\right),$$



$$P\{\bar{X} < \mu\} = \Phi\left(\frac{\mu - \mu}{\sigma/\sqrt{10}}\right) = \Phi(0) = \frac{1}{2}.$$

(2) 若一人得奖的概率为  $p$ , 则得奖人数  $Y \sim b(10, p)$ . 此处  $p$  是随机选取一人, 其考分  $X$  在 70 分以上的概率. 因  $X \sim N(62, 25)$ , 故

$$\begin{aligned} p &= P\{X > 70\} = 1 - P\{X \leq 70\} = 1 - \Phi\left(\frac{70-62}{\sqrt{25}}\right) \\ &= 1 - \Phi(1.6) = 1 - 0.9452 = 0.0548. \end{aligned}$$

至少一人得奖的概率为

$$P\{Y \geq 1\} = 1 - (0.9452)^{10} = 0.431.$$

6. 设总体  $X \sim b(1, p)$ ,  $X_1, X_2, \dots, X_n$  是来自  $X$  的样本.

(1) 求  $(X_1, X_2, \dots, X_n)$  的分布律.

(2) 求  $\sum_{i=1}^n X_i$  的分布律.

(3) 求  $E(\bar{X}), D(\bar{X}), E(S^2)$ .

解 (1) 因  $X_1, X_2, \dots, X_n$  相互独立, 且有  $X_i \sim b(1, p), i=1, 2, \dots, n$ , 即  $X_i$  具有分布律  $P\{X_i = x_i\} = p^{x_i}(1-p)^{1-x_i}, x_i=0, 1$ , 因此  $(X_1, X_2, \dots, X_n)$  的分布律为

$$\begin{aligned} P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ &= \prod_{i=1}^n P\{X_i = x_i\} \\ &= \prod_{i=1}^n [p^{x_i}(1-p)^{1-x_i}] = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

(2) 因  $X_1, X_2, \dots, X_n$  相互独立, 且有  $X_i \sim b(1, p), i=1, 2, \dots, n$ , 故  $\sum_{i=1}^n X_i \sim b(n, p)$ , 其分布律为

$$P\left\{\sum_{i=1}^n X_i = k\right\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k=0, 1, 2, \dots, n.$$

(3) 由于总体  $X \sim b(1, p), E(X)=p, D(X)=p(1-p)$ , 故有

$$E(\bar{X}) = p, D(\bar{X}) = \frac{p(1-p)}{n}, E(S^2) = D(X) = p(1-p).$$

7. 设总体  $X \sim \chi^2(n), X_1, X_2, \dots, X_{10}$  是来自  $X$  的样本, 求  $E(\bar{X}), D(\bar{X}), E(S^2)$ .

解 因  $E(X)=n, D(X)=2n$ , 故有

$$E(\bar{X}) = n, \quad D(\bar{X}) = \frac{2n}{10} = \frac{n}{5},$$

而

$$E(S^2) = D(X) = 2n.$$



8. 设总体  $X \sim N(\mu, \sigma^2)$ ,  $X_1, X_2, \dots, X_{10}$  是来自  $X$  的样本.

(1) 写出  $X_1, X_2, \dots, X_{10}$  的联合概率密度.

(2) 写出  $\bar{X}$  的概率密度.

解 (1) 由假设  $X_i (i=1, 2, \dots, 10)$  的概率密度为

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i - \mu)^2 / (2\sigma^2)},$$

故  $X_1, X_2, \dots, X_{10}$  的联合概率密度为

$$\begin{aligned} \prod_{i=1}^{10} f_{X_i}(x_i) &= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i - \mu)^2 / (2\sigma^2)} \\ &= \frac{1}{(2\pi\sigma^2)^5} e^{-\sum_{i=1}^{10} (x_i - \mu)^2 / (2\sigma^2)}. \end{aligned}$$

(2)  $\bar{X} \sim N(\mu, \frac{\sigma^2}{10})$ , 故  $\bar{X}$  的概率密度为

$$f_{\bar{X}}(x) = \frac{\sqrt{10}}{\sqrt{2\pi}\sigma} e^{-5(x - \mu)^2 / \sigma^2} = \frac{\sqrt{5}}{\sqrt{\pi}\sigma} e^{-5(x - \mu)^2 / \sigma^2}.$$

9. 设在总体  $N(\mu, \sigma^2)$  中抽得一容量为 16 的样本, 这里  $\mu, \sigma^2$  均未知.

(1) 求  $P\{S^2/\sigma^2 \leq 2.041\}$ , 其中  $S^2$  为样本方差.

(2) 求  $D(S^2)$ .

解 (1) 因为  $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ , 现在  $n=16$ , 即有  $\frac{15S^2}{\sigma^2} \sim \chi^2(15)$ , 故有

$$\begin{aligned} p &= P\{S^2/\sigma^2 \leq 2.041\} = P\{15S^2/\sigma^2 \leq 15 \times 2.041\} \\ &= P\{15S^2/\sigma^2 \leq 30.615\} = 1 - P\{15S^2/\sigma^2 > 30.615\}. \end{aligned}$$

查  $\chi^2$  分布表得  $\chi_{0.01}^2(15) = 30.578$ , 从而知  $p = 1 - 0.01 = 0.99$ .

(2) 由  $15S^2/\sigma^2 \sim \chi^2(15)$ , 得

$$D(15S^2/\sigma^2) = 2 \times 15 = 30,$$

即

$$\frac{15^2}{\sigma^4} D(S^2) = 30, \quad D(S^2) = \frac{2\sigma^4}{15}.$$

10. 下面列出了 30 个美国 NBA 球员的体重(以磅计, 1 磅 = 0.454 kg.) 数据. 这些数据是从美国 NBA 球队 1990—1991 赛季的花名册中抽样得到的.

225	232	232	245	235	245	270	225	240	240
217	195	225	185	200	220	200	210	271	240
220	230	215	252	225	220	206	185	227	236

(1) 画出这些数据的频率直方图(提示: 最大和最小观察值分别为 271 和 185, 区间  $[184.5, 271.5]$  包含所有数据, 将整个区间分为 5 等份, 为计算方便, 将



区间调整为(179.5, 279.5).

(2) 作出这些数据的箱线图.

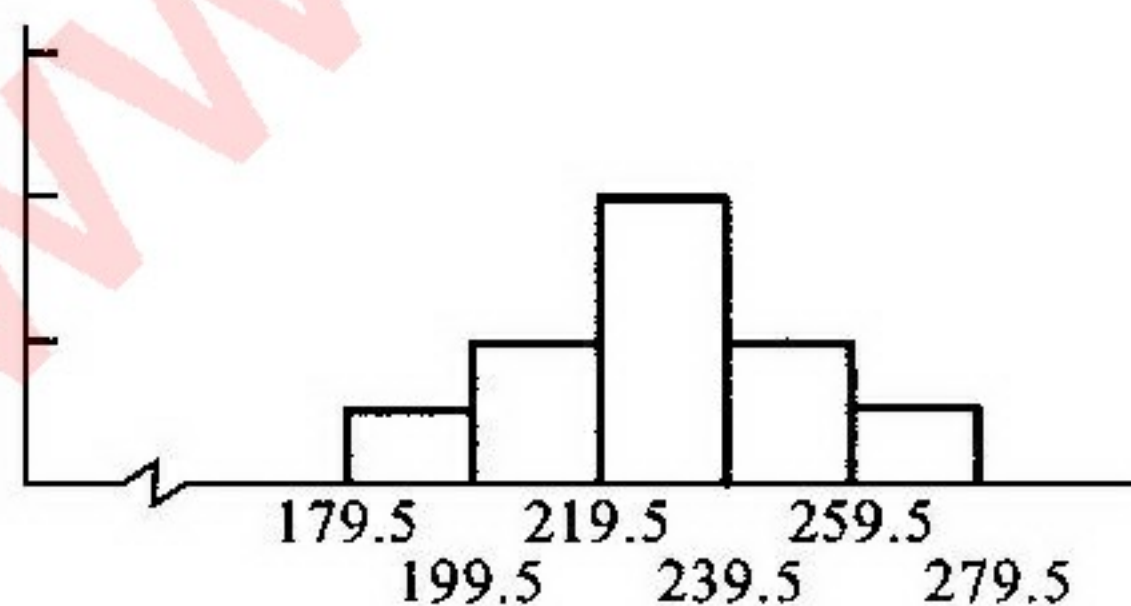
解 (1) 最大和最小观察值分别为 271 和 185, 考虑到这些数据是将实测数据经四舍五入后得到的, 取区间  $I=[184.5, 271.5]$  使得所有实测数据都落在  $I$  上. 将区间  $I$  等分为若干小区间, 小区间的个数与数据个数  $n$  有关, 取为  $\sqrt{n}$  左右为佳. 现在取小区间的个数为 5, 于是小区间的长度为  $(271.5 - 184.5)/5 = 17.4$ . 这一长度使用起来不方便. 为此, 将区间  $I$  的下限延伸至 179.5, 上限延伸至 279.5. 这样小区间的长度调整为

$$\Delta = (279.5 - 179.5)/5 = 20.$$

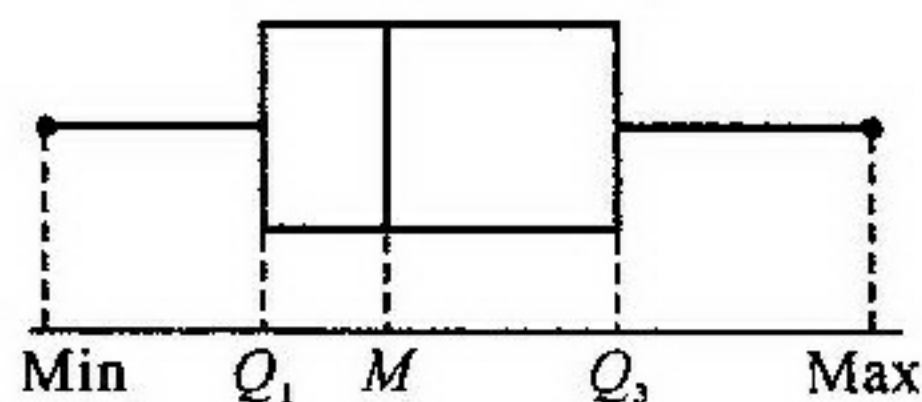
数出落在每个小区间内的数据的个数  $f_i, i=1, 2, 3, 4, 5$ , 算出数据落在各个小区间的频率  $f_i/n (n=30, i=1, 2, 3, 4, 5)$ , 所得结果列表如下:

组限	频数 $f_i$	频率 $f_i/n$	累积频率
179.5~199.5	3	0.1	0.10
199.5~219.5	6	0.2	0.30
219.5~239.5	13	0.43	0.73
239.5~259.5	6	0.2	0.93
259.5~279.5	2	0.07	1

在每个小区间上作以对应的频率除以  $\Delta$  为高(即以  $(f_i/n)/\Delta$  为高)以小区间为底的小长方形. 小长方形的面积就是  $[(f_i/n)/\Delta] \times \Delta = f_i/n$ . 画出图形, 这就是所求的频率直方图(如题 6.10 图(1)).



(1)



(2)

题 6.10 图

(2) 将  $n=30$  个数据按自小到大的次序排序得到

185 185 195 200 200 206 210 215 217 220  
 220 220 225 225 225 225 227 230 232 232  
 235 236 240 240 240 245 245 252 270 271

下面来求第一四分位数  $Q_1$ , 中位数  $M$ , 第三四分位数  $Q_3$ .



因  $np = 30 \times 0.25 = 7.5$ , 故  $Q_1$  位于左起第  $[7.5] + 1 = 8$  处, 即有  $Q_1 = 215$ .

因  $np = 30 \times 0.5 = 15$ , 故  $M = Q_2$  是这 30 个数最中间两个数的平均值, 即有  $Q_2 = M = \frac{1}{2}(225 + 225) = 225$ .

因  $np = 30 \times 0.75 = 22.5$ , 故  $Q_3$  位于左起第  $[22.5] + 1 = 23$  处, 即有  $Q_3 = 240$ . 又  $\text{Min} = 185, \text{Max} = 271$ .

根据  $\text{Min}, Q_1, M, Q_3, \text{Max}$  这 5 点作出箱线图如题 6.10 图(2)所示. 从上述两个图能看出数据的分布关于中心线比较对称.

**11. 截尾均值** 设数据集包含  $n$  个数, 将这些数据从小到大排序为

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)},$$

删去  $100\alpha\%$  个数值小的数, 同时删去  $100\alpha\%$  个数值大的数, 将留下的数据取算术平均, 记为  $\bar{x}_\alpha$ , 即

$$\bar{x}_\alpha = \frac{x_{([n\alpha]+1)} + \cdots + x_{(n-[n\alpha])}}{n - 2[n\alpha]}$$

其中,  $[n\alpha]$  是小于或等于  $n\alpha$  的最大整数 (一般取  $\alpha = 0.1 \sim 0.2$ ).  $\bar{x}_\alpha$  称为  $100\alpha\%$  截尾均值. 例如对于第 10 题中的 30 个数, 取  $\alpha = 0.1$ , 则有  $[n\alpha] = [30 \times 0.1] = 3$ , 得  $100 \times 0.1\%$  截尾均值为

$$\bar{x}_\alpha = \frac{200 + 200 + \cdots + 245 + 245}{30 - 6} = 225.4167.$$

若数据来自某一总体的样本, 则  $\bar{x}_\alpha$  是一个统计量.  $\bar{x}_\alpha$  不受样本的极端值的影响. 截尾均值在实际应用问题中是常会用到的.

试求第 10 题的数据的  $\alpha = 0.2$  的截尾值.

**解**  $\alpha = 0.2, [n\alpha] = [30 \times 0.2] = 6, 100 \times 0.2\%$  截尾均值为

$$\bar{x}_\alpha = \frac{210 + 215 + \cdots + 240 + 240}{30 - 12} = 226.3333.$$