鹿児島大学　渡邉研究室
wata104@eee

中央値の分散：管理図係数 m_3 の計算式

1．中央値

　n 個のデータ x_i（$i = 1, 2, \cdots, n$）を大きさの順に並べたとき（小さい順でも大きい順でもよい）、n が奇数の場合はちょうどその中央に当たる 1 つの値、n が偶数の場合は中央の 2 つの値の算術平均を、中央値（メディアン）という。すなわち、中央値 $\tilde{x}$ は

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & n\text{ が奇数の場合} \\ \dfrac{x_{n/2} + x_{(n/2)+1}}{2}, & n\text{ が偶数の場合} \end{cases}$$

で表される。

　中央値 $\tilde{x}$ は、n 個のデータ x_i について、β を基準値とする偏差の絶対値の和

$$S(\beta) = \sum_{i=1}^{n} |x_i - \beta|$$

を最小にする β に相当する。これは以下のように示される。

$$S(\beta) = \sum_{i=1}^{n} |x_i - \beta| = -\sum_{x_i<\beta} (x_i - \beta) + \sum_{x_i>\beta} (x_i - \beta)$$

より、$x_i < \beta$ となる x_i の個数を n_-、$x_i > \beta$ となる x_i の個数を n_+ とすると

$$\frac{\partial S(\beta)}{\partial \beta} = -\sum_{x_i<\beta} (-1) + \sum_{x_i>\beta} (-1) = n_- - n_+$$

である。n が奇数の場合は、x_i を大きさの順に並べたとき

$$\beta = x_{(n+1)/2}$$

とすると $n_- = n_+$ となり、$\partial S(\beta)/\partial \beta = 0$ より $S(\beta)$ が最小となる。また、n が偶数の場合は、x_i を小さい順（昇順）に並べたとき

$$x_{n/2} < \beta < x_{(n/2)+1} \qquad \text{if} \quad x_{n/2} \neq x_{(n/2)+1}$$
$$x_{n/2} = \beta = x_{(n/2)+1} \qquad \text{if} \quad x_{n/2} = x_{(n/2)+1}$$

とすれば $n_- = n_+$ となり、$\partial S(\beta)/\partial \beta = 0$ より $S(\beta)$ が最小となる。$x_{n/2} \neq x_{(n/2)+1}$ のとき、β は一意に定まらないが、

$$\beta = \frac{x_{n/2} + x_{(n/2)+1}}{2}$$

とすれば、$x_{n/2} = x_{(n/2)+1}$ のときも含めて $\partial S(\beta)/\partial \beta = 0$ となる。

　なお、算術平均は、n 個のデータ x_i について、α を基準値とする偏差の 2 乗の和

$$S(\alpha) = \sum_{i=1}^{n} (x_i - \alpha)^2$$

を最小にする α に等しい（付録 1 を参照）。

2. 確率密度関数と累積分布関数

区間 $-\infty < x < \infty$ で定義される確率変数が $x \sim x + dx$ の間の値をとる確率を確率密度関数 $f(x)$ で表す。累積分布関数 $F(x)$ は、確率変数が x 以下の値をとる確率であり、

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

で表される。ここで、$F(-\infty) = 0,\ F(\infty) = 1$ である。また、

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

が成り立つ。

3. 中央値の期待値

確率密度関数 $f(x)$ が平均値

$$\bar{x} = \int_{-\infty}^{\infty} xf(x)dx$$

を中心に対称な分布であるときは、中央値 $\tilde{x}$ は平均値 $\bar{x}$ に一致する。特に、確率密度関数 $f(x)$ が標準正規分布

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

にしたがうとき、平均値 $\bar{x}$ と中央値 $\tilde{x}$ はともに 0 である。

確率密度関数 $f(x)$ が非対称な分布であるときは、次節に示す中央値の 2 乗の期待値と同様にして、中央値の期待値を求めることができる。

4. 中央値の 2 乗の期待値

n 個のデータ x_i（$i = 1, 2, \cdots, n$）の中央値 $\tilde{x}$ の分散を求めるために、まず、中央値 $\tilde{x}$ の 2 乗の期待値 $\overline{\tilde{x}^2}$ を求める。

n が奇数の場合、ある x_i に対して、それとは異なる $(n-1)/2$ 個の x_j が $x_j \le x_i$ になり、それ以外の $(n-1)/2$ 個の x_k が $x_i \le x_k$ になる確率を考えて

$$\begin{aligned}
\overline{\tilde{x}^2} &= \sum_{i=1}^{n} {}_{n-1}C_{(n-1)/2} \int_{-\infty}^{\infty} {x_i}^2 f(x_i) F(x_i)^{(n-1)/2} (1 - F(x_i))^{(n-1)/2} dx_i \\
&= n \cdot {}_{n-1}C_{(n-1)/2} \int_{-\infty}^{\infty} x^2 F'(x) F(x)^{(n-1)/2} (1 - F(x))^{(n-1)/2} dx \\
&= n \frac{(n-1)!}{\left(\frac{n-1}{2}\right)! \left((n-1) - \frac{n-1}{2}\right)!} \int_{-\infty}^{\infty} x^2 F'(x) F(x)^{(n-1)/2} (1 - F(x))^{(n-1)/2} dx \\
&= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \int_{-\infty}^{\infty} x^2 F'(x) F(x)^{(n-1)/2} (1 - F(x))^{(n-1)/2} dx
\end{aligned}$$

$$= \frac{n!}{(m!)^2}\int_{-\infty}^{\infty} x^2 F'(x)F(x)^m(1-F(x))^m dx$$

となる。ただし、$m = (n-1)/2$ である。

n が偶数の場合は、ある x_i に対して、それとは異なる x_j（$j \neq i$）が $x_i \leq x_j$ になり、残りの $n-2$ 個の x_k（$k \neq i, j$）のうち、$(n-2)/2$ 個の x_k が $x_k \leq x_i$ になり、それ以外の $(n-2)/2$ 個の x_k が $x_j \leq x_k$ になる確率を考えて

$$\overline{\tilde{x}^2} = \sum_{i=1}^{n}\sum_{j\neq i}^{n} {}_{n-2}C_{(n-2)/2}$$

$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{x_j}\left(\frac{x_i+x_j}{2}\right)^2 f(x_i)f(x_j)F(x_i)^{(n-2)/2}\left(1-F(x_j)\right)^{(n-2)/2} dx_i\,dx_j$$

$$= n(n-1)_{n-2}C_{(n-2)/2}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\left(\frac{x+y}{2}\right)^2 F'(x)F'(y)F(x)^{(n-2)/2}(1-F(y))^{(n-2)/2} dx\,dy$$

となる。ここで、

$$n(n-1)_{n-2}C_{(n-2)/2} = n(n-1)\frac{(n-2)!}{\left(\frac{n-2}{2}\right)!\left((n-2)-\frac{n-2}{2}\right)!} = \frac{n!}{\left\{\left(\frac{n-2}{2}\right)!\right\}^2}$$

であるから

$$\overline{\tilde{x}^2} = \frac{n!}{\left\{\left(\frac{n-2}{2}\right)!\right\}^2}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\left(\frac{x+y}{2}\right)^2 F'(x)F'(y)F(x)^{(n-2)/2}(1-F(y))^{(n-2)/2} dx\,dy$$

$$= \frac{n!}{(m!)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\left(\frac{x+y}{2}\right)^2 F'(x)F'(y)F(x)^m(1-F(y))^m dx\,dy$$

となる。ただし、$m = (n-2)/2$ である。

5. 中央値の分散と標準偏差

n 個のデータ x_i の中央値 $\tilde{x}$ の分散 $V(\tilde{x})$ は

$$V(\tilde{x}) = \overline{\tilde{x}^2} - \bar{\tilde{x}}^2$$

により求められ、中央値 $\tilde{x}$ の標準偏差は

$$D(\tilde{x}) = \sqrt{V(\tilde{x})} = \sqrt{\overline{\tilde{x}^2} - \bar{\tilde{x}}^2}$$

により求められる。

確率変数 x が元の変数 X の平均 μ と標準偏差 σ によって $x = (X-\mu)/\sigma$ と正規化された値であるとき、X の中央値 Me の標準偏差 $D(Me)$ と x の中央値 $\tilde{x}$ の標準偏差 $D(\tilde{x})$ との関係は

$$D(Me) = \sigma D(\tilde{x})$$

となる。いっぽう、n 個のデータ X_i が独立であるとき、その平均値 $\bar{X}$ の標準偏差は

$$D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

である（付録２を参照）。したがって、$D(Me)$ と $D(\bar{X})$ の比を、データの個数 n によって決まる係数 m_3 で表して

$$D(Me) = m_3 D(\bar{X}) = m_3 \frac{\sigma}{\sqrt{n}}$$

とすると

$$m_3 = \frac{\sqrt{n}}{\sigma} D(Me) = \sqrt{n} D(\tilde{x}) = \sqrt{n(\overline{\tilde{x}^2} - \bar{\tilde{x}}^2)}$$

となる。$\bar{\tilde{x}} = 0$ のときは $m_3 = \sqrt{n\overline{\tilde{x}^2}}$ より ${m_3}^2 = n\overline{\tilde{x}^2}$ であるから、n が奇数の場合は、$m = (n-1)/2$ として

$${m_3}^2 = \frac{n \cdot (n!)}{(m!)^2} \int_{-\infty}^{\infty} x^2 F'(x) F(x)^m (1 - F(x))^m dx$$

n が偶数の場合は、$m = (n-2)/2$ として

$${m_3}^2 = \frac{n \cdot (n!)}{(m!)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{y} \left(\frac{x+y}{2}\right)^2 F'(y) F'(x) F(x)^m (1 - F(y))^m dx\, dy$$

である。

累積分布関数 $F(x)$ が標準正規分布

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$$

にしたがうとき、上式の $-\infty$ から $+\infty$ までの積分をシンプソン法により-5 から$+5$ まで刻み幅 0.01 で数値積分して m_3 の値を計算すると、下の表のようになる。

表　m_3、d_2、A_4 の値

n	m_3	d_2	A_4
2	1.0000	1.12838	1.8800
3	1.1602	1.69257	1.1872
4	1.0922	2.05875	0.7957
5	1.1976	2.32593	0.6908
6	1.1351	2.53441	0.5485
7	1.2137	2.70436	0.5089
8	1.1600	2.84720	0.4321
9	1.2227	2.97003	0.4117
10	1.1762	3.07750	0.3626

上の表の d_2 は、n 個のデータ X_i の範囲（最大値と最小値の差）の平均 $\bar{R}$ と標準偏差

σ の比を $\bar{R} = d_2\sigma$ として表したときの係数である（その計算方法は、別資料「範囲の期待値：係数 d_2 の計算式」を参照）。また、A_4 は

$$A_4 = m_3 \frac{3}{d_2\sqrt{n}}$$

で表される値であり、品質管理で用いられる $Me - R$ 管理図で、Me の管理限界を

$$E(Me) \pm 3D(Me) = \overline{Me} \pm m_3 \frac{3\sigma}{\sqrt{n}} = \overline{Me} \pm m_3 \frac{3}{d_2\sqrt{n}}\bar{R} = \overline{Me} \pm m_3 A_2 \bar{R} = \overline{Me} \pm A_4 \bar{R}$$

として定める際に用いられる。ここで、A_2 は、$\bar{X} - R$ 管理図で、$\bar{X}$ の管理限界を

$$E(\bar{X}) \pm 3D(\bar{X}) = \bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}} = \bar{\bar{X}} \pm \frac{3}{d_2\sqrt{n}}\bar{R} = \bar{\bar{X}} \pm A_2\bar{R}$$

として定める際に用いられる係数であり、$A_4 = m_3 A_2$ である。

参考文献

・森口 繁一「初等数理統計学」改訂版、培風館、1957
・中村 達男「管理図の作り方と活用」（新版 QC 入門講座 7）、日本規格協会、1999
・奥村 士郎「品質管理入門テキスト」改訂 2 版、日本規格協会、2007
・稲本 稔、細野 泰彦「わかりやすい品質管理」第 4 版、オーム社、2016

（付録１）

n 個のデータ x_i（$i = 1, 2, \cdots, n$）について、α を基準値とする偏差の二乗の和

$$S(\alpha) = \sum_{i=1}^{n} (x_i - \alpha)^2$$

を最小にする α は

$$\frac{\partial S(\alpha)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (x_i - \alpha)^2 = -2 \sum_{i=1}^{n} (x_i - \alpha) = -2 \left(\sum_{i=1}^{n} x_i - n\alpha \right) = 0$$

より

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} x_i$$

となる。これは x_i の算術平均である。

（付録２）

母平均 μ、母標準偏差 σ の母集団からサンプリングした大きさ n のサンプル x_i（$i = 1, 2, \cdots, n$）の平均 $\bar{x}$ の分散の期待値は

$$E[V(\bar{x})] = E[(\bar{x}-\mu)^2] = E\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu\right)^2\right] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)\right)^2\right]$$

$$= \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n}(x_i - \mu)\right)^2\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(x_i - \mu)^2\right] + \frac{1}{n^2}E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}(x_i - \mu)(x_j - \mu)\right]$$

となる。ここで、x_i が独立であるとき、

$$E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}(x_i - \mu)(x_j - \mu)\right] = 0$$

であるから

$$E[V(\bar{x})] = \frac{1}{n^2}E\left[\sum_{i=1}^{n}(x_i - \mu)^2\right] = \frac{1}{n}E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right] = \frac{\sigma^2}{n}$$

となる。したがって、大きさ n のサンプル x_i（$i = 1, 2, \cdots, n$）の平均 $\bar{x}$ の標準偏差は

$$D(\bar{x}) = \sqrt{E[V(\bar{x})]} = \frac{\sigma}{\sqrt{n}}$$

である。