鹿児島大学　渡邉研究室
wata104@eee

範囲の期待値と分散：管理図係数 d_2, d_3 の計算式

1. 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。

2. 範囲の期待値

確率変数 x_i のサンプルの最大値 x_L と最小値 x_S の差を範囲（レンジ）という。大きさ n のサンプルの $x_i\,(i = 1, 2, \cdots, n)$ が各々独立に確率密度関数 $f(x)$ の分布にしたがうとき、x_i の範囲 r の期待値は、ある x_i に対して、それとは異なる $x_j\,(j \neq i)$ が $x_j \leq x_i$ になり、x_i と x_j 以外の $n-2$ 個の x_k （$k \neq i, j$） が $x_j \leq x_k \leq x_i$ になる確率を考えて

$$\bar{r} = \overline{x_L - x_S} = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} (x_i - x_j) f(x_i) f(x_j) \big(F(x_i) - F(x_j)\big)^{n-2} dx_j\, dx_i$$

$$= n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y - x) f(y) f(x) (F(y) - F(x))^{n-2} dx\, dy$$

$$= n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y - x) F'(y) F'(x) (F(y) - F(x))^{n-2} dx\, dy$$

$$= n \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y - x) \cdot (n-1) F'(x) (F(y) - F(x))^{n-2} dx\, F'(y) dy$$

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

となる。ここで、

$$-\frac{\partial}{\partial x} \left((F(y) - F(x))^{n-1}\right) = \frac{\partial}{\partial x} \left(F(y)^{n-1} - (F(y) - F(x))^{n-1}\right)$$

を用いると

$$-\int_{-\infty}^{y} (y - x) \cdot \frac{\partial}{\partial x} \left((F(y) - F(x))^{n-1}\right) dx$$

$$= \int_{-\infty}^{y} (y - x) \cdot \frac{\partial}{\partial x} \left(F(y)^{n-1} - (F(y) - F(x))^{n-1}\right) dx$$

$$= \left[(y-x)\cdot\left(F(y)^{n-1} - (F(y)-F(x))^{n-1}\right) \ \right]_{x=-\infty}^{x=y} + \int_{-\infty}^{y} \left(F(y)^{n-1} - (F(y)-F(x))^{n-1}\right) dx$$

$$= \int_{-\infty}^{y} \left(F(y)^{n-1} - (F(y)-F(x))^{n-1}\right) dx$$

であるから

$$\bar{r} = n \int_{-\infty}^{\infty} \int_{-\infty}^{y} \left(F(y)^{n-1} - (F(y)-F(x))^{n-1}\right) dx\, F'(y) dy$$

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

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y} \frac{\partial}{\partial y} \left(F(y)^{n} - (F(y)-F(x))^{n}\right) dx\, dy$$

となり、x と y の積分の順序を変更すると

$$\bar{r} = \int_{-\infty}^{\infty} \int_{x}^{\infty} \frac{\partial}{\partial y} \left(F(y)^{n} - (F(y)-F(x))^{n}\right) dy\, dx$$

$$= \int_{-\infty}^{\infty} \left[F(y)^{n} - (F(y)-F(x))^{n} \ \right]_{y=x}^{y=\infty} dx$$

$$= \int_{-\infty}^{\infty} \left(1 - (1-F(x))^{n} - F(x)^{n}\right) dx$$

$$= \int_{-\infty}^{\infty} \left(1 - F(x)^{n} - (1-F(x))^{n}\right) dx$$

が得られる。なお、この式は、最大値 x_L と最小値 x_S の期待値の差から導くこともできる（別資料「範囲の期待値：管理図係数 d_2 の計算式」を参照）。

3. 範囲の 2 乗の期待値

大きさ n のサンプル x_i（$i = 1, 2, \cdots, n$）の範囲 r の分散を求めるために、まず、範囲 r の 2 乗の期待値を求める。ある x_i に対して、それとは異なる x_j（$j \neq i$）が $x_j \leq x_i$ になり、x_i と x_j 以外の $n-2$ 個の x_k（$k \neq i, j$）が $x_j \leq x_k \leq x_i$ になる確率を考えて

$$\overline{r^2} = \overline{(x_L - x_S)^2} = \sum_{i=1}^{n} \sum_{j\neq i}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} \left(x_i - x_j\right)^2 f(x_i) f(x_j) \left(F(x_i) - F(x_j)\right)^{n-2} dx_j\, dx_i$$

$$= n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y-x)^2 f(y) f(x) (F(y)-F(x))^{n-2} dx\, dy$$

$$= n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y-x)^2 F'(y) F'(x) (F(y)-F(x))^{n-2} dx\, dy$$

$$= n \int_{-\infty}^{\infty} \int_{-\infty}^{y} (y-x)^2 \cdot (n-1) F'(x) (F(y)-F(x))^{n-2} dx\, F'(y) dy$$

鹿児島大学　渡邉研究室
wata104@eee

$$= -n\int_{-\infty}^{\infty}\int_{-\infty}^{y}(y-x)^2\cdot\frac{\partial}{\partial x}((F(y)-F(x))^{n-1})dx\,F'(y)dy$$

となる。ここで、

$$-\frac{\partial}{\partial x}((F(y)-F(x))^{n-1}) = \frac{\partial}{\partial x}(F(y)^{n-1}-(F(y)-F(x))^{n-1})$$

を用いると

$$-\int_{-\infty}^{y}(y-x)^2\cdot\frac{\partial}{\partial x}((F(y)-F(x))^{n-1})dx$$

$$= \int_{-\infty}^{y}(y-x)^2\cdot\frac{\partial}{\partial x}(F(y)^{n-1}-(F(y)-F(x))^{n-1})dx$$

$$= \left[(y-x)^2\cdot(F(y)^{n-1}-(F(y)-F(x))^{n-1})\ \right]_{x=-\infty}^{x=y}$$

$$+\int_{-\infty}^{y}2(y-x)\cdot(F(y)^{n-1}-(F(y)-F(x))^{n-1})dx$$

$$= 2\int_{-\infty}^{y}(y-x)\cdot(F(y)^{n-1}-(F(y)-F(x))^{n-1})dx$$

であるから

$$\overline{r^2} = 2n\int_{-\infty}^{\infty}\int_{-\infty}^{y}(y-x)\cdot(F(y)^{n-1}-(F(y)-F(x))^{n-1})dx\,F'(y)dy$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{y}(y-x)\cdot n(F(y)^{n-1}-(F(y)-F(x))^{n-1})F'(y)dx\,dy$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{y}(y-x)\cdot\frac{\partial}{\partial y}(F(y)^{n}-(F(y)-F(x))^{n})dx\,dy$$

となり、x と y の積分の順序を変更すると

$$\overline{r^2} = 2\int_{-\infty}^{\infty}\int_{x}^{\infty}(y-x)\cdot\frac{\partial}{\partial y}(F(y)^{n}-(F(y)-F(x))^{n})dy\,dx$$

となる。さらに、

$$\frac{\partial}{\partial y}(F(y)^{n}-(F(y)-F(x))^{n}) = \frac{\partial}{\partial y}(-1+F(y)^{n}+(1-F(x))^{n}-(F(y)-F(x))^{n})$$

を用いると

$$\int_{x}^{\infty}(y-x)\cdot\frac{\partial}{\partial y}(F(y)^{n}-(F(y)-F(x))^{n})dy$$

$$= \int_{x}^{\infty}(y-x)\cdot\frac{\partial}{\partial y}(-1+F(y)^{n}+(1-F(x))^{n}-(F(y)-F(x))^{n})dy$$

$$= \left[(y-x)\cdot(-1+F(y)^{n}+(1-F(x))^{n}-(F(y)-F(x))^{n})\ \right]_{y=x}^{y=\infty}$$

$$+\int_{x}^{\infty}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dy$$

$$=\int_{x}^{\infty}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dy$$

であるから

$$\overline{r^2}=2\int_{-\infty}^{\infty}\int_{x}^{\infty}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dy\,dx$$

となり、再び x と y の積分の順序を変更すると

$$\overline{r^2}=2\int_{-\infty}^{\infty}\int_{-\infty}^{y}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dx\,dy$$

が得られる。

4. 範囲の分散

大きさ n のサンプル x_i の範囲 r の期待値は

$$\bar{r}=\int_{-\infty}^{\infty}(1-F(x)^n-(1-F(x))^n)dx$$

により求められ、範囲 r の 2 乗の期待値は

$$\overline{r^2}=2\int_{-\infty}^{\infty}\int_{-\infty}^{y}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dx\,dy$$

により求められるから、範囲 r の分散 $V(r)$ は

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

$$=2\int_{-\infty}^{\infty}\int_{-\infty}^{y}(1-F(y)^n-(1-F(x))^n+(F(y)-F(x))^n)dx\,dy$$

$$-\left(\int_{-\infty}^{\infty}(1-F(x)^n-(1-F(x))^n)dx\right)^2$$

により求められる。

5. 範囲と標準偏差

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 の範囲 R の期待値 $\bar{R}$ と標準偏差 σ の関係は

$$\bar{R}=\int_{-\infty}^{\infty}(1-F(X)^n-(1-F(X))^n)dX$$

$$=\sigma\int_{-\infty}^{\infty}(1-F(x)^n-(1-F(x))^n)dx$$

となる。したがって、$\bar{R}$ と σ の比をサンプルの大きさ n によって決まる係数 d_2 で表して $\bar{R}=d_2\sigma$ とすると

$$d_2 = \frac{\bar{R}}{\sigma} = \int_{-\infty}^{\infty} (1 - F(x)^n - (1 - F(x))^n) dx$$

となる。

また、R の 2 乗の期待値 $\overline{R^2}$ と標準偏差 σ の関係は

$$\overline{R^2} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{Y} (1 - F(Y)^n - (1 - F(X))^n + (F(Y) - F(X))^n) dX\,dY$$

$$= 2\sigma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{y} (1 - F(y)^n - (1 - F(x))^n + (F(y) - F(x))^n) dx\,dy$$

となる。したがって、R の分散 $V(R)$ は

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

$$= 2\sigma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{y} (1 - F(y)^n - (1 - F(x))^n + (F(y) - F(x))^n) dx\,dy - {d_2}^2\sigma^2$$

となり、R の標準偏差 $D(R) = \sqrt{V(R)}$ と σ の比をサンプルの大きさ n によって決まる係数 d_3 で表して $D(R) = d_3\sigma$ とすると

$$ {d_3}^2 = \frac{V(R)}{\sigma^2} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{y} (1 - F(y)^n - (1 - F(x))^n + (F(y) - F(x))^n) dx\,dy - {d_2}^2$$

となる。

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 d_2 と d_3 の値を計算すると、下の表のようになる。

表　d_2 と d_3 の値

n	d_2	d_3
2	1.12838	0.85251
3	1.69257	0.88838
4	2.05875	0.87982
5	2.32593	0.86409
6	2.53441	0.84805
7	2.70436	0.83322
8	2.84720	0.81984
9	2.97003	0.80784
10	3.07750	0.79706

この d_2 と d_3 の値は、品質管理で用いられる $\bar{X} - R$ 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、また、R の管理限界を

$$E(R) + 3D(R) = \bar{R} + 3d_3\sigma = \bar{R} + \frac{3d_3}{d_2}\bar{R} = \left(1 + \frac{3d_3}{d_2}\right)\bar{R} = D_4\bar{R}$$

$$E(R) - 3D(R) = \bar{R} - 3d_3\sigma = \bar{R} - \frac{3d_3}{d_2}\bar{R} = \left(1 - \frac{3d_3}{d_2}\right)\bar{R} = D_3\bar{R}$$

として定める際に用いられる。

参考文献

・統計科学研究会「新編 統計数値表」河出書房、1952
・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