スピアマンの順位相関係数の式の導出

スピアマンの順位相関係数の計算方法は各 x_i，y_i が順位であること以外はピアソンの積率相関係数（単に相関係数と呼ばれることが多い）と同一である。なお，添え字の無い $\sum$ は $\sum_{i=1}^{n}$ を表すものとする。

$$\rho_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

S_{xx}，S_{yy}，S_{xy} はそれぞれ x の偏差平方和，y の偏差平方和，x と y の偏差積和といい，定義式から以下のように変形される。偏差平方和，偏差積和はデータ数（またはデータ数 -1）で割ると分散，共分散になるため，相関係数は共分散をそれぞれの標準偏差の積で割ったものともいえる。

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}，\ S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}，\ S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

1.同順位(タイ)を含まない場合

各 x_i，y_i は 1 から n までの整数になるから，その総和や平方和は以下のように表される。

$$\sum x_i = \sum y_i = \sum i = \frac{n(n+1)}{2}，\ \sum x_i^2 = \sum y_i^2 = \sum i^2 = \frac{n(n+1)(2n+1)}{6}$$

これらを S_{xx} に代入して以下のように変形する。

$$\begin{aligned} S_{xx} &= \sum x_i^2 - \frac{(\sum x_i)^2}{n} = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \\ &= \frac{2n(n+1)(2n+1) - 3n(n+1)^2}{12} = \frac{n(n+1)\{2(2n+1) - 3(n+1)\}}{12} \\ &= \frac{n(n+1)(n-1)}{12} \end{aligned}$$

同様にして，$S_{yy} = \dfrac{n(n+1)(n-1)}{12}$ も求められる。

x と y の積和については以下のように x と y の差の平方和で表す式に変形する。

$$\begin{aligned} \sum x_i y_i &= \frac{2\sum x_i y_i}{2} = \frac{2\sum x_i y_i - \sum x_i^2 - \sum y_i^2 + \sum x_i^2 + \sum y_i^2}{2} \\ &= \frac{\sum x_i^2 + \sum y_i^2 - \sum(x_i^2 - 2x_i y_i + y_i^2)}{2} \\ &= \frac{2\sum i^2 - \sum(x_i - y_i)^2}{2} = \sum i^2 - \frac{\sum(x_i - y_i)^2}{2} \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{\sum(x_i - y_i)^2}{2} \end{aligned}$$

これを S_{xy} に代入して以下のように変形する。

$$\begin{aligned} S_{xy} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{\sum(x_i - y_i)^2}{2} - \frac{(\sum i)^2}{n} \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} - \frac{\sum(x_i - y_i)^2}{2} \\ &= \frac{n(n+1)(n-1)}{12} - \frac{\sum(x_i - y_i)^2}{2} \end{aligned}$$

以上より，同順位を含まない場合のスピアマンの順位相関係数の式が求められる。

$$\rho_{xy}=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}=\frac{S_{xy}}{\sqrt{S_{xx}S_{xx}}}=S_{xy}\cdot\frac{1}{S_{xx}}$$
$$=\left\{\frac{n(n+1)(n-1)}{12}-\frac{\sum(x_i-y_i)^2}{2}\right\}\frac{12}{n(n+1)(n-1)}$$
$$=1-\frac{6\sum(x_i-y_i)^2}{n(n+1)(n-1)}$$

2.同順位(タイ)を含む場合

同順位のデータは順位の平均値をそのデータの順位とする。例えば1位タイが3つ存在する場合は3つとも2位とし，5位タイが2つ存在する場合は2つとも5.5位とする。よって，x_i，y_iの総和は同順位を含まない場合と変わらないが，その平方和やxとyの積和は同順位を含まない場合とは異なる。

x_iの中にm_x位タイがt_x個存在する場合，すなわちm_x位から(m_x+t_x-1)位までが同記録である場合のx_iの平方和は，1からnまでの整数の平方和から同順位の部分を引き，代わりに順位の平均値の平方を加えるため，以下のように表される。

$$\sum x_i^2=\sum i^2-\sum_{j=m_x}^{m_x+t_x-1}j^2+t_x\left(m_x+\frac{t_x-1}{2}\right)^2$$
$$=\frac{n(n+1)(2n+1)}{6}-\sum_{j=1}^{m_x+t_x-1}j^2+\sum_{j=1}^{m_x-1}j^2+t_x\left(m_x+\frac{t_x-1}{2}\right)^2$$
$$=\frac{n(n+1)(2n+1)}{6}-\frac{(m_x+t_x-1)(m_x+t_x)(2m_x+2t_x-1)}{6}+\frac{(m_x-1)m_x(2m_x-1)}{6}+t_x\left(m_x+\frac{t_x-1}{2}\right)^2$$
$$=\frac{n(n+1)(2n+1)}{6}+\frac{-t_x^3+t_x}{12}$$
$$=\frac{n(n+1)(2n+1)}{6}-\frac{t_x(t_x+1)(t_x-1)}{12}$$

このことから，必要なのは同順位がいくつ存在するかだけであり，どの順位が同順位であるかは計算結果に影響を及ぼさないことが分かる。上記の式は同順位が1箇所のみの場合であるが，同順位がn_x箇所存在する場合はj箇所目の同順位数を$t_{x(j)}$個$(j=1,\ 2,\ \cdots,\ n_x)$として，x_iの平方和は以下のように表される。

$$\sum x_i^2=\frac{n(n+1)(2n+1)}{6}-\frac{\sum_{j=1}^{n_x}t_{x(j)}(t_{x(j)}+1)(t_{x(j)}-1)}{12}$$

これをS_{xx}に代入して以下のように変形する。

$$S_{xx}=\sum x_i^2-\frac{(\sum x_i)^2}{n}$$
$$=\frac{n(n+1)(2n+1)}{6}-\frac{\sum_{j=1}^{n_x}t_{x(j)}(t_{x(j)}+1)(t_{x(j)}-1)}{12}-\frac{n(n+1)^2}{4}$$
$$=\frac{n(n+1)(n-1)-\sum_{j=1}^{n_x}t_{x(j)}(t_{x(j)}+1)(t_{x(j)}-1)}{12}$$

同様にして，y_iの中に同順位がn_y箇所存在しk箇所目の同順位数が$t_{y(k)}$個$(k=1,\ 2,\ \cdots,\ n_y)$である場合のS_{yy}も求められる。

$$S_{yy}=\frac{n(n+1)(n-1)-\sum_{k=1}^{n_y}t_{y(k)}(t_{y(k)}+1)(t_{y(k)}-1)}{12}$$

x と y の積和については以下のように x と y の差の平方和と S_{xx}，S_{yy} で表す形に変形する。

$$
\begin{aligned}
\sum x_i y_i &= \frac{\sum x_i^2 + \sum y_i^2 - \sum (x_i - y_i)^2}{2} \\
&= \frac{S_{xx} + (\sum x_i)^2/n + S_{yy} + (\sum y_i)^2/n - \sum (x_i - y_i)^2}{2} \\
&= \frac{S_{xx} + n(n+1)^2/4 + S_{yy} + n(n+1)^2/4 - \sum (x_i - y_i)^2}{2} \\
&= \frac{S_{xx} + S_{yy} - \sum (x_i - y_i)^2}{2} + \frac{n(n+1)^2}{4}
\end{aligned}
$$

これを S_{xy} に代入して以下のように変形する。

$$
\begin{aligned}
S_{xy} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \\
&= \frac{S_{xx} + S_{yy} - \sum (x_i - y_i)^2}{2} + \frac{n(n+1)^2}{4} - \frac{n(n+1)^2}{4} \\
&= \frac{S_{xx} + S_{yy} - \sum (x_i - y_i)^2}{2}
\end{aligned}
$$

以上より，x_i の中に同順位が n_x 箇所存在し j 箇所目の同順位数が $t_{x(j)}$ 個($j = 1$，2，…，n_x)，y_i の中に同順位が n_y 箇所存在し k 箇所目の同順位数が $t_{y(k)}$ 個($k = 1$，2，…，n_y)である場合のスピアマンの順位相関係数の式が求められる。

$$
\begin{aligned}
\rho_{xy} &= \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{S_{xx} + S_{yy} - \sum (x_i - y_i)^2}{2\sqrt{S_{xx} S_{yy}}} \\
S_{xx} &= \frac{n(n+1)(n-1) - \sum_{j=1}^{n_x} t_{x(j)}(t_{x(j)}+1)(t_{x(j)}-1)}{12} \\
S_{yy} &= \frac{n(n+1)(n-1) - \sum_{k=1}^{n_y} t_{y(k)}(t_{y(k)}+1)(t_{y(k)}-1)}{12}
\end{aligned}
$$

【例①】M-1 グランプリ 2008 の中田カウスの採点と合計点の順位相関係数

表 1：M-1 グランプリ 2008 のファーストラウンド採点結果(抜粋)

コンビ名	中田カウス	合計点	x_i：カウス順位	y_i：合計点順位	$(x_i - y_i)^2$
オードリー	98	649	1	1	0
NON STYLE	91	644	5	2	9
ナイツ	94	640	2	3	1
笑い飯	88	637	7	4	9
U 字工事	92	623	4	5	1
ダイアン	93	619	3	6	9
モンスターエンジン	90	614	6	7	1
キングコング	86	612	8	8	0
ザ・パンチ	85	591	9	9	0

同順位は存在せず，$n = 9$，$\sum (x_i - y_i)^2 = 30$ である。

$$
\therefore \rho_{xy} = 1 - \frac{6\sum (x_i - y_i)^2}{n(n+1)(n-1)} = 1 - \frac{6 \cdot 30}{9 \cdot 10 \cdot 8} = 0.75
$$

【例②】M-1 グランプリ 2002 の立川談志の採点と合計点の順位相関係数

表 2：M-1 グランプリ 2002 のファーストラウンド採点結果(抜粋)

コンビ名	立川談志	合計点	x_i：談志順位	y_i：合計点順位	$(x_i - y_i)^2$
フットボールアワー	70	621	5.5	1	20.25
ますだおかだ	80	612	1.5	2	0.25
笑い飯	70	567	5.5	3	6.25
おぎやはぎ	80	561	1.5	4	6.25
ハリガネロック	70	545	5.5	5	0.25
テツ and トモ	70	539	5.5	6	0.25
スピードワゴン	50	535	9	7	4
ダイノジ	70	534	5.5	8	6.25
アメリカザリガニ	70	525	5.5	9	12.25

x_i に同順位が存在し，$n = 9$，$\sum(x_i - y_i)^2 = 56$，$n_x = 2$，$t_{x(1)} = 2$，$t_{x(2)} = 6$ である。

$$S_{xx} = \frac{n(n+1)(n-1) - \sum_{j=1}^{n_x} t_{x(j)}(t_{x(j)}+1)(t_{x(j)}-1)}{12} = \frac{9 \cdot 10 \cdot 8 - (2 \cdot 3 \cdot 1 + 6 \cdot 7 \cdot 5)}{12} = 42$$

$$S_{yy} = \frac{n(n+1)(n-1)}{12} = \frac{9 \cdot 10 \cdot 8}{12} = 60$$

$$\therefore \rho_{xy} = \frac{S_{xx} + S_{yy} - \sum(x_i - y_i)^2}{2\sqrt{S_{xx} S_{yy}}} = \frac{42 + 60 - 56}{2\sqrt{42 \cdot 60}} = 0.4582$$