□ Pearson correlation coefficient :

$$r_{xy} = \frac{s_{xy}}{s_x \times s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

□ correlation coefficient in Q-methodology :

$$r_{xy} = 1 - \frac{\sum d^2}{2ns^2} \text{ where } d = x - y, \quad n: \text{ number of Q-sorts}, \quad s = s_x = s_y : stdev$$

□ Spearman's rank correlation coefficient :

$$r_{xy} = 1 - \frac{6\sum d^2}{n(n^2-1)} \text{ where } d = x - y, \quad n: \text{ number of observations}$$

1. 요인분석 중 Q-type(Q-methodology)에 사용되는 상관계수 $r_{xy} = 1 - \frac{\sum d^2}{2ns^2}$ 는 아래에서 처럼 Pearson 상관계수에서 수학적으로 도출된다. 단, Q 방법론에서는 점수를 강제로 표준화하므로 모든 Q-sort의 평균과 표준편차가 동일하다.

$$\begin{aligned}
r_{xy} &= \frac{s_{xy}}{s_x \times s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \\
&= \frac{\sum (x_i - m)(y_i - m)}{ns^2} \quad (\because \bar{x} = \bar{y} = m, \sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})^2 = ns^2) \\
&= \frac{\sum (x_i y_i - m x_i - m y_i + m^2)}{ns^2} = \frac{\sum x_i y_i - nm^2}{ns^2} \quad (\because \sum x_i = \sum y_i = nm) \\
&= \frac{\sum (x_i - y_i)^2 - \sum (x_i - y_i)^2 + 2\sum x_i y_i - 2nm^2}{2ns^2} = \frac{\sum x_i^2 + \sum y_i^2 - 2nm^2 - \sum (x_i - y_i)^2}{2ns^2} \\
&= \frac{-\sum (x_i - y_i)^2 + 2ns^2}{2ns^2} \quad (\because Var(X) = E(X^2) - E(X)^2 \Rightarrow \sum x_i^2 - nm^2 = \sum y_i^2 - nm^2 = ns^2) \\
&= 1 - \frac{\sum d_i^2}{2ns^2} \quad where\ d_i = x_i - y_i
\end{aligned}$$

2. 또한, Q 방법론의 상관계수는 “서열척도”의 Spearman의 순위상관계수 $r_{xy} = 1 - \frac{6\sum d^2}{n(n^2-1)}$ 와 동일하다. 단, Q 방법론은 순위상관계수와 달리 순위의 중복을 허용한 경우이다(예를 들어, 2위가 다수 존재할 수 있다). 만약 Q 방법론에서 순위 중복을 허용하지 않는다면(즉, 피라미드가 아니라 일렬 배치만 허용한다면) 다음과 같다.

$$
\begin{aligned}
s^2 &= \frac{\sum (x - \frac{n+1}{2})^2}{n} \quad (\because \bar{x} = \frac{n+1}{2}) \\
&= \frac{\sum (x^2 - (n+1)x + \frac{(n+1)^2}{4})}{n} \\
&= \frac{\sum x^2 - (n+1)\frac{n(n+1)}{2} + \frac{n(n+1)^2}{4}}{n} \quad (\because \sum x = \frac{n(n+1)}{2}) \\
&= \frac{\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}}{n} \quad (\because \sum x^2 = \frac{n(n+1)(2n+1)}{6}) \\
&= \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} = \frac{n^2 - 1}{12}
\end{aligned}
$$

그러므로 $r_{xy} = 1 - \frac{\sum d_i^2}{2ns^2} = 1 - \frac{\sum d_i^2}{2n(\frac{n^2-1}{12})} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$ 가 된다.

3. 즉, Spearman의 순위상관계수는 값의 중복이 없는 표준화된(최대값과 최소값이 강제된) 데이터(순위/서열척도)의 Pearson 상관계수일 뿐이고, Q 방법론은 값의 중복이 있되 표준화된(최대값과 최소값 및 중복의 개수가 강제된) 데이터(순위/서열척도)의 Pearson 상관계수일 뿐이다. 결국, 서열척도든 Q 방법론이든 Pearson 상관계수로 유사성을 계산한다고 보면 된다(그러므로 우리는 서열척도든 Q 방법론이든 Pearson 상관계수를 사용한다고 이해하면 된다).

   다만, 참고로 d_i^2은 거리개념이므로, 표준화된(최대값과 최소값이 강제된) 데이터는 거리개념(군집분석)을 사용하던 상관개념(요인분석)을 사용하던 유사할 것이라는 부가적인 정보는 얻을 수 있다.