Equivalence of Pearson Correlation for Ranks and Spearman's ρ

Charles Nicholson
cnicholson@ou.edu

August 2015

Let $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$. Let $\bar{x}$ and $\bar{y}$ denote the mean of X and Y, respectively. Pearson's correlation coefficient is:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Assume that X and Y are ranks. For simplicity, assume there are no ties in ranks. Since there are no ties, both variables consist of the integers from 1 to n inclusive.

And we can rewrite the denominator:

$$\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \tag{1}$$

Recall the following:

$$\sum_i^k = \frac{(k)(k+1)}{2}$$

$$\sum_i^k i^2 = \frac{(k)(k+1)(2k+1)}{6}$$

Therefore,

$$\sum_i x_i = \frac{(n)(n+1)}{2}$$

$$\sum_i x_i^2 = \frac{(n)(n+1)(2n+1)}{6}$$

$$\bar{x} = \frac{\sum_i x_i}{n} = \frac{n+1}{2}$$

Also note,

$$\bar{x} = \bar{y}$$

$$\sum_i x_i \bar{x} = \bar{x} \sum_i x_i = \frac{1}{n} \sum_i x_i \sum_i x_i = n\bar{x}^2$$

Now the denominator in (1) is just a function of n:

$$\begin{aligned}
\sum_i (x_i - \bar{x})^2 &= \sum_i x_i^2 - \sum_i 2x_i\bar{x} + \sum_i \bar{x} \\
&= \sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_i x_i^2 - n\bar{x}^2 \\
&= \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2 \\
&= n(n+1)\left(\frac{2n+1}{6} - \frac{n+1}{4}\right) \\
&= n(n+1)\left(\frac{8n+4-6n-6}{24}\right) \\
&= n(n+1)\left(\frac{n-1}{12}\right) \\
&= \frac{n(n^2-1)}{12}
\end{aligned}$$

The numerator from (1):

$$\begin{aligned}
\sum_i (x_i - \bar{x})(y_i - \bar{y}) &= \sum_i x_i(y_i - \bar{y}) - \sum_i \bar{x}(y_i - \bar{y}) \\
&= \sum_i x_i y_i - \bar{y}\sum_i x_i - \bar{x}\sum_i y_i + n\bar{x}\bar{y} \\
&= \sum_i x_i y_i - n\bar{x}\bar{y} \\
&= \sum_i x_i y_i - n\left(\frac{n+1}{2}\right)^2 \\
&= \sum_i x_i y_i - \frac{n(n+1)}{12} 3(n+1) \\
&= \frac{n(n+1)}{12}(-3(n+1)) + \sum_i x_i y_i \\
&= \frac{n(n+1)}{12}\left[(n-1) - (4n+2)\right] + \sum_i x_i y_i \\
&= \frac{n(n+1)(n-1)}{12} - \frac{n(n+1)(2n+1)}{6} + \sum_i x_i y_i \\
&= \frac{n(n+1)(n-1)}{12} - \sum_i x_i^2 + \sum_i x_i y_i \\
&= \frac{n(n+1)(n-1)}{12} - \sum_i \frac{x_i^2 + y_i^2}{2} + \sum_i x_i y_i \\
&= \frac{n(n+1)(n-1)}{12} - \sum_i \frac{x_i^2 - 2x_i y_i + y_i^2}{2} \\
&= \frac{n(n+1)(n-1)}{12} - \sum_i \frac{(x_i - y_i)^2}{2}
\end{aligned}$$

Putting these results together as numerator/denominator:

$$
\begin{aligned}
&= \frac{\dfrac{n(n+1)(n-1)}{12} - \sum_i \dfrac{(x_i - y_i)^2}{2}}{\dfrac{n(n^2-1)}{12}} \\
&= \frac{\dfrac{n(n^2-1)}{12} - \sum_i \dfrac{(x_i - y_i)^2}{2}}{\dfrac{n(n^2-1)}{12}} \\
&= 1 - \frac{6\sum_i (x_i - y_i)^2}{n(n^2-1)}
\end{aligned}
$$

Hence,

$$\rho = 1 - \frac{6\sum_i (x_i - y_i)^2}{n(n^2-1)}$$