Correlation Q2 [Problem/H](8/6/21)

Show that the formula $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ can be written as

$r_s = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)}$ when the data items are ranks.

[In other words, instead of using the formula for Spearman's coefficient, it is theoretically possible to use the standard formula for r.]

Show that the formula $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ can be written as

$r_s = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)}$ when the data items are ranks.

[In other words, instead of using the formula for Spearman's coefficient, it is theoretically possible to use the standard formula for r.]

Solution

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

As the x_i are just the numbers 1 to n in some order,

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

and so $S_{xx} = \frac{n}{6}(n+1)(2n+1) - \frac{n(n+1)^2}{4}$

By the same reasoning, S_{yy} will also have this value,

so the denominator of r is

$\frac{n}{6}(n+1)(2n+1) - \frac{n(n+1)^2}{4} = \frac{n}{12}(n+1)\{4n+2-(3n+3)\}$

$= \frac{n}{12}(n+1)(n-1)$

Then $S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

Now $\sum {d_i}^2 = \sum (x_i - y_i)^2 = (\sum {x_i}^2) + (\sum {y_i}^2) - 2\sum x_i y_i$,

so that $\sum x_i y_i = \frac{(\sum {x_i}^2) + (\sum {y_i}^2) - \sum {d_i}^2}{2}$

$= \frac{1}{2}\left\{2.\frac{n}{6}(n+1)(2n+1) - \sum {d_i}^2\right\}$

Hence $S_{xy} = \frac{n}{6}(n+1)(2n+1) - \frac{1}{2}\sum {d_i}^2 - \frac{\left(\frac{n}{2}(n+1)\right)^2}{n}$

$= \frac{n}{12}(n+1)\{4n+2-(3n+3)\} - \frac{1}{2}\sum {d_i}^2$

$$= \frac{n}{12}(n+1)(n-1) - \frac{1}{2}\sum {d_i}^2$$

and $r = \frac{\frac{n}{12}(n+1)(n-1) - \frac{1}{2}\sum {d_i}^2}{\frac{n}{12}(n+1)(n-1)} = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)}$