Rank Correlation (5 pages; 23/8/19)

(1) Formula

(1.1) The x and y coordinates of each data point (of n points) are replaced by their x and y ranks (where the smallest coordinate can be given either rank 1 or n, provided that the other coordinate is ranked in the same order). If the (x_i, y_i) are now pairs of ranks, and $d_i = x_i - y_i$, then

Spearman's coefficient of rank correlation (r_s) is defined as

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

If the plotted ranks give rise to a strictly increasing or decreasing curve, then r_s will be 1 or –1.

(1.2) In fact, the Spearman formula $r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$ is just a rearrangement of the Pearson formula $\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ when the data items are ranks, and so it is possible (though usually more complicated) to use the Pearson formula instead.

However, because in the case of ranked data there is no assumption of a bivariate Normal distribution, different tables apply.

Proof

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

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

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

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

(2) Use of the rank correlation coefficient

(2.1) It can be applied when the underlying data doesn't have a bivariate Normal distribution.

(2.2) The association of the underlying data need not be linear, but it should be a 'monotonic relationship'; ie increasing or decreasing.

(2.3) For some data, only ranks may be available (or they are easier to obtain); eg the positions awarded by judges in a competition.

(2.4) If the underlying data is available, and the conditions for PMCC are satisfied, then the PMCC provides a better test, as otherwise information is lost in converting to ranks.

(2.5) The formula isn't strictly applicable if any ranks are tied; however, it will be approximately correct if only a few ranks are tied.

(3) Hypothesis Tests

(3.1) Critical values for r_s are slightly different from those for r (because of the different assumptions regarding the bivariate Normal distribution).

(3.2) Critical Values for r_s:

	5%	2½%	1%	½%
	10%	5%	2%	1%
n				
1	–	–	–	–
2	–	–	–	–
3	–	–	–	–
4	1.0000	–	–	–
5	0.9000	1.0000	1.0000	–
6	0.8286	0.8857	0.9429	1.0000
7	0.7143	0.7857	0.8929	0.9286
8	0.6429	0.7381	0.8333	0.8810
9	0.6000	0.7000	0.7833	0.8333
10	0.5636	0.6485	0.7455	0.7939
11	0.5364	0.6182	0.7091	0.7545
12	0.5035	0.5874	0.6783	0.7273
13	0.4835	0.5604	0.6484	0.7033
14	0.4637	0.5385	0.6264	0.6791
15	0.4464	0.5214	0.6036	0.6536
16	0.4294	0.5029	0.5824	0.6353
17	0.4142	0.4877	0.5662	0.6176
18	0.4014	0.4716	0.5501	0.5996
19	0.3912	0.4596	0.5351	0.5842
20	0.3805	0.4466	0.5218	0.5699

Critical Values for r, for comparison:

	5%	2½%	1%	½%
	10%	5%	2%	1%
n				
1	–	–	–	–
2	–	–	–	–
3	0.9877	0.9969	0.9995	0.9999
4	0.9000	0.9500	0.9800	0.9900
5	0.8054	0.8783	0.9343	0.9587
6	0.7293	0.8114	0.8822	0.9172
7	0.6694	0.7545	0.8329	0.8745
8	0.6215	0.7067	0.7887	0.8343
9	0.5822	0.6664	0.7498	0.7977
10	0.5494	0.6319	0.7155	0.7646
11	0.5214	0.6021	0.6851	0.7348
12	0.4973	0.5760	0.6581	0.7079
13	0.4762	0.5529	0.6339	0.6835
14	0.4575	0.5324	0.6120	0.6614
15	0.4409	0.5140	0.5923	0.6411
16	0.4259	0.4973	0.5742	0.6226
17	0.4124	0.4821	0.5577	0.6055
18	0.4000	0.4683	0.5425	0.5897
19	0.3887	0.4555	0.5285	0.5751
20	0.3783	0.4438	0.5155	0.5614