Rank correlation coefficient

- Since the assumptions underlying the significance test for the correlation coefficient are rather stringent, it is sometimes preferable to use a non-parametric alternative.
- Most popular among the non-parametric measures of association is the rank correlation coefficient also called Spearman's rank correlation coefficient.
- For a given set of data $(x_1, y_1), (x_2, y_2), (x_3, y_n) \cdots (x_n, y_n)$,
- The rank correlation is obtained by ranking the x_i's among themselves and the y's both from low to high or high to low, the rank correlation coefficient is given by:

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

where d_i is the difference ~~between~~ between the ranks assigned to x_i and y_i. Where there are ties in the ranks we assigned the tied observations the mean of the ranks which they jointly occupy.

Examples

Calculate the Spearman's coefficient rank correlation from the following.

x	73	118	115	107	95	81	79	75
y	147	125	132	137	130	140	139	125

Calculate r.

x	r_x	y	r_y	δ_i	δ_i^2
73	1	147	8	−7	49
118	8	125	1	7	49
115	7	132	3	4	16
101	6	137	4	2	4
95	5	130	2	3	9
81	4	140	6	−2	4
79	3	139	5	−2	4
75	2	145	7	−5	25

$\Sigma \delta_i^2 = 160$

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} \delta_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(160)}{8(63)}$$

$$= 1 - \frac{120}{63}$$

$$= -0.9048$$

- For small values of n, i.e $n \leq 10$ the test of H_0 of no correlation may be based on special tables determined from the exact sampling distribution of r_s.

- Most of the time ~~through~~ though we use the fact that the distribution of random variables can be a closely approximated to the normal distribution

Theorem: Under the null hypothesis of no correlation the mean and variance of random variables are:

$E[r_s = 0]$ and $Var(r_s) = 1/_{n-1}$

Note: Strictly speaking the theorem applies when there are no ties, but the result can be used as an approximation unless the number of ties is large

Example

1) For the data in the previous example test at $\alpha = 0.09$ whether the rank correlation coefficient is significant

Solution

We want to test

H_0: There is no correlation

H_1: There is correlation

The test statistic is;

$$Z = \frac{r_s - E[r_s]}{\sqrt{Var(r_s)}} = \frac{r_s - 0}{\sqrt{1/_{n-1}}} = r_s\sqrt{n-1}$$

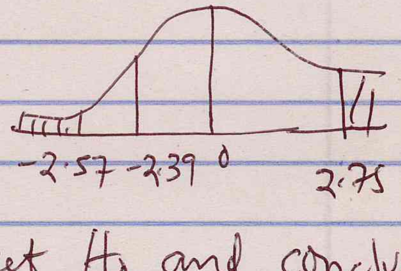
We reject Ho if ~~Ro~~ $Z \leq -2.75$ or $Z \geq 2.75$

i.e $Z_{\alpha/2} = Z_{0.09/2} = Z_{0.005} = \pm 2.75$

$r_s = -0.9048$, $n = 8$

$\Rightarrow Z = -0.9048\sqrt{8-1}$

$= -2.3938$



$Z = -2.3938 > -2.575$ We do reject H_0 and conclude that the correlation is not significant

Note: When there are no ties in ranks, r_s actually equals the correlation coefficient r, calculated for the ranks.

Let r_i and s_i be the ranks of x_i and y_i respectively.

- We find that

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = \frac{n(n+1)}{2} \quad (1+2+3+\cdots+n)$$

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} s_i^2 = \frac{n(n+1)(2n+1)}{6} \quad (1^2+2^2+\cdots+n^2)$$

Let $d_i = r_i - s_i$

$$d_i^2 = (r_i - s_i)^2 = r_i^2 - 2r_i s_i + s_i^2$$

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} r_i^2 - 2\sum_{i=1}^{n} r_i s_i + \sum_{i=1}^{n} s_i^2$$

$$\Rightarrow 2\sum_{i=1}^{n} r_i s_i = \sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} s_i^2 - \sum_{i=1}^{n} d_i^2$$

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

$$r = \frac{\sum r_i s_i - \frac{1}{n}\left(\sum r_i\right)\left(\sum s_i\right)}{\sqrt{\sum\left(r_i^2 - n\bar{r}^2\right)\sum\left(s_i^2 - n\bar{s}^2\right)}}$$

$$= \frac{n(n+1)(2n+1) - \frac{1}{2}\Sigma d_i^2 - \frac{1}{n}\left(\frac{n(n+1)}{2}\right)\left(\frac{n(n+1)}{2}\right)}{\sqrt{\left[\left(\frac{n(n+1)(2n+1)}{6}\right) - \frac{n(n+1)^2}{2}\right]\left[\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2}\right]^2}}$$

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

Note: $\Sigma r_i^2 - n\bar{r}^2$, $\bar{r} = \frac{1}{n}\Sigma r_i^2$

$$= \frac{1}{n}\left(\frac{n(n+1)}{2}\right)$$

$$n\bar{r}^2 = n\left(\frac{1}{n^2}\left(\frac{n(n+1)}{2}\right)^2\right.$$

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

$$= 1 - \frac{\frac{1}{2}\sum_{i=1}^{n} d_i^2}{\frac{n(n+1)(2n+1)}{6} - \frac{1}{n}\left(\frac{n(n+1)}{2}\right)^2} \quad -- (*)$$

$$\frac{n(n+1)(2n+1)}{6} - \frac{1}{n}\left(n\left(\frac{n+1}{2}\right)\right)^2 = \frac{(n+1)}{12}\left[2n(2n+1) - 3n(n+1)\right]$$

$$= \frac{n+1}{12}\left[4n^2 - 2n - 3n^2 - 3n\right]$$

$$= \frac{n+1}{12}\left(n^2 - n\right)$$

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

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

Replacing in eqn (*) we find that

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

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

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

check expansion of

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

$$\frac{1}{6}\left\{n(2n^2+3n+1)\right\} = \frac{1}{6}\left\{2n^3+3n^2+n\right\}$$

7

$$\frac{1}{n}\left(\frac{n^2+n}{2}\right)^2 = \frac{1}{4n}\left(n^4+2n^3+n^2\right)$$

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

$\Rightarrow$

$$\frac{2n^3+3n^2+n}{6} - \left(\frac{n^3+2n^2}{4}+n\right)$$

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

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