Different ranking approaches defining association and agreement measures of paired ordinal data

Elisabeth Svensson
Örebro University, Sweden

ISCB30 PRAGUE

Subjective ratings

Observer rating, clinical judgements

level of impairment, severity of disease
diagnostic certainty, clinical status
social functioning, activity of daily life

Self-assessment

QoL, quality of health, sleep disturbance
stress, anxiety

Generating ordinal data

Verbal descriptive scales (VDS)

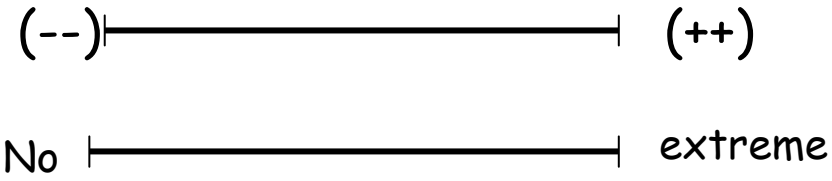
How much….?
[ ] Extremely high
[ ] Very high
[ ] Moderate
[ ] Slight
[ ] Very low
[ ] non-existing

How often….?
[ ] None of the time
[ ] A little of the time
[ ] Some of the time
[ ] Most of the time
[ ] All of the time

Numeric Rating Scales (NRS)

0 1 2 3 4

Visual Analogue Scales (VAS)



Pictogram

- Hierarchical Conditional Scales (Staircases)
- Categorised Continuous Quantitative Data

Rank-invariant properties of ordinaldata

Ordered categorical data are **invariant** under all order-preserving transformations

Non-numbers: Any set of ordered symbols can be used (1,2,3,.., a,b,c…, +, ++, +++, ..)

Change in labelling must not affect the result of data analysis

NO ARITHMETICS (Sums, differences).

AIM

To compare the applications of two ranking approaches to paired ordinal data

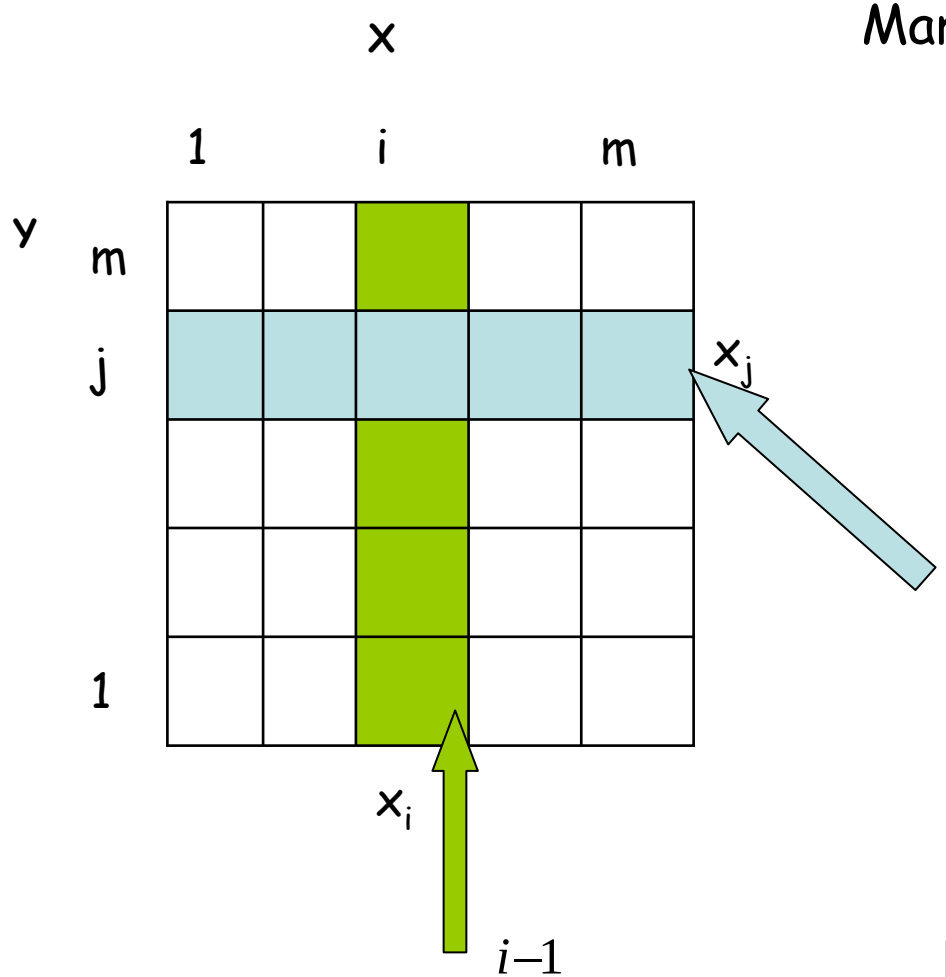
Association:

- variables X: $\{1,\ldots,i,\ldots,m_1\}$, Y:$\{1,\ldots,j,\ldots,m_2\}$

Agreement:

-assessments X:$\{1,\ldots,i,\ldots,m\}$, Y:$\{1,\ldots,j,\ldots,m\}$

Classical ranking, paired data



Marginal distributions $\{x_i\}$, $\{x_j\}$,

i, j: 1,…,m

Ranks tied to each marginal

$$\bar{r}_j = \sum_{\nu=1}^{j-1} x_\nu + \tfrac{1}{2}(x_j + 1)$$

$$\bar{r}_i = \sum_{\nu=1}^{i-1} x_\nu + \tfrac{1}{2}(x_i + 1)$$

Association

The Spearman rank-order correlation coefficient – a measure of **association** between variables

based on rank-differences

adjusted for ties $t^{(X)}$, $t^{(Y)}$

$$d_\nu = r_\nu^{(X)} - r_\nu^{(Y)}$$

$$r_s = \frac{n^3 - n - 6\sum_{\nu=1}^{m} d_\nu^2 - \tfrac{1}{2}(t^{(X)} + t^{(Y)})}{\sqrt{\left(n^3 - n\right)^2 - (t^{(X)} + t^{(Y)})(n^3 - n) + t^{(X)}t^{(Y)}}}$$

without ties:

$$r_s = 1 - \frac{6\sum_{k=1}^{n} d_k^2}{n^3 - n} \qquad k=1, \ldots, n$$

Relationship between perceived back pain and physical health?

		How severe back **pain** have you had during the past four weeks?					
		None	Negli-gible	Mode-rate	Rather severe	Very severe	Tot
Regarding your back pain, how is your per-ceived **physical health?**	Very bad		**1**	**2**	**6**	**4**	13
	Rather bad		**5**	**13**	**24**	**3**	45
	Neither ..nor		**2**	**12**	**8**	**2**	24
	Rather good	**1**		**10**	**6**	**1**	18 **rank 10.5**
	Very good			**1**			1
	Total	1	8	38 **Rank 28.5**	44	10	101

r_s 0.24

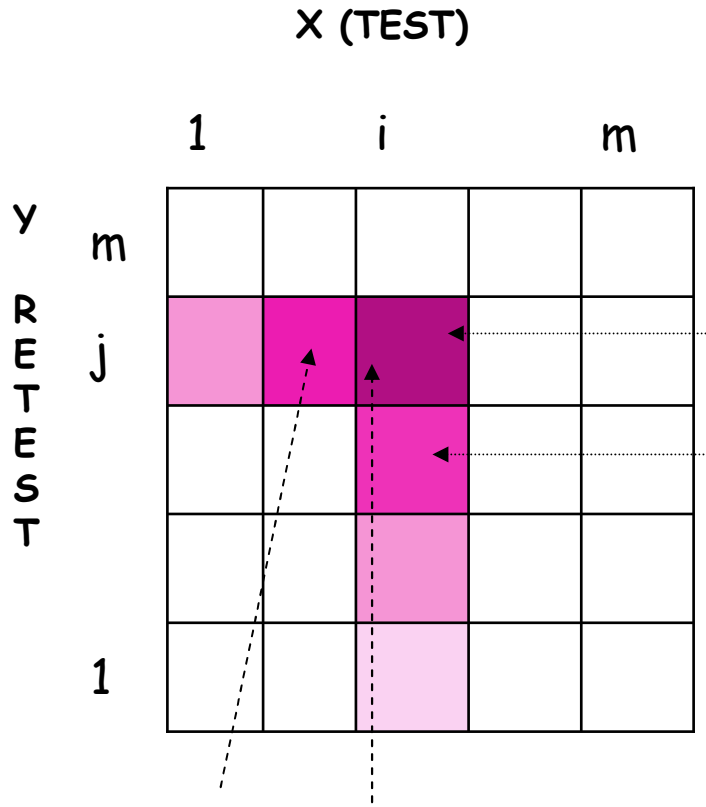
Agreement

RELIABILITY:
The extent to which repeated assessments yield the same result ➔

AGREEMENT
Inter- rater agreement
intra-rater agreement
stability

Augmented ranks based on information from pairs of assessments

The augmented ranking of paired ordinal data



Aug-rank: The ranks are tied to the cells

$$\overline{R}^{(X)}_{i(j-1)} < \overline{R}^{(X)}_{ij}$$

$$\overline{R}^{(Y)}_{(i-1)j} < \overline{R}^{(Y)}_{ij}$$

Agreement

(Svensson) Method based on aug-ranks

Developed for paired ordinal data from all types of scales (incl VAS)

Identify and measure the components of

- **systematic** and
- **individual**

sources of disagreement

Intra-rater agreement in test-retest assessments of perceived back pain

How severe low back pain have you had during the last four weeks?

Percentage Agreement (PA)

78/101= 77%

		First occasion: None	Negligible	Moderate	Rather severe	Very severe	
Second occasion	Very severe				3	3	6
	Rather severe			5	36	7	48
	Moderate		2	33	5		40
	Negligible		5				5
	None	1	1				2
		1	8	38	44	10	101

The pairs of aug-ranks

X: First occasion

Y	None	Negli-gible	Mode-rate	Rather severe	Very severe
Very severe				3 **(90;97)**	3 **(100;100)**
Rather severe			5 **(45;50)**	36 **(70.5;70.5)**	7 **(95;92)**
Mode-rate		2 **(8.5;8.5)**	33 **(26;26)**	5 **(50;45)**	
Negli-gible		5 **(5;5)**			
None	1 **(1;1)**	1 **(2;2)**			

$$\overline{R}_{21} < \overline{R}_{22} < \overline{R}_{23}$$

The pairs of aug-ranks

X: First occasion

Y	None	Negli-gible	Mode-rate	Rather severe	Very severe
Very severe				3 **(90;97)**	3 **(100;100)**
Rather severe			5 **(45;50)**	36 **(70.5;70.5)**	7 **(95;92)**
Mode-rate		2 **(8.5;8.5)**	33 **(26;26)**	5 **(50;45)**	
Negli-gible		5 **(5;5)**			
None	1 **(1;1)**	1 **(2;2)**			

$$\overline{R}_{ij}^{(X)} \neq \overline{R}_{ij}^{(Y)}$$

Dispersion from the Rank-Transformable Pattern of Agreement (RTPA)

The rank transformable pattern of agreement, RTPA

- pairing off the two marginal distributions
- the expected pattern when systematic disagreement only

$$\overline{R}_{ij}^{(X)} = \overline{R}_{ij}^{(Y)}$$

Y \ X	None	Negli-gible	Mode-rate	Rather severe	Very severe	Total
Very severe					6	6
Rather severe				44	4	48
Moderate		2	38			40
Negligible		5				5
None	1	1				2
Total	1	8	38	44	10	101

The Relative Rank Variance, RV, the measure of

- dispersed pairs from the RTPA
- individual/occasional disagreement ("noise")
- Heterogeneity of the group

Y \ X	None	Negli-gible	Mode-rate	Rather severe	Very severe
Very severe				3 **(90;97)**	3 **(100;100)**
Rather severe			5 **(45;50)**	36 **(70.5;70.5)**	7 **(95;92)**
Mode-rate		2 **(8.5;8.5)**	33 **(26;26)**	5 **(50;45)**	
Negli-gible		5 **(5;5)**			
None	1 **(1;1)**	1 **(2;2)**			

$$\Delta\overline{R}_{ij} = \overline{R}_{ij}^{(X)} - \overline{R}_{ij}^{(Y)}$$

$$RV = \frac{6}{n^3}\sum_{i=1}^{m}\sum_{j=1}^{m} x_{ij}\Delta\overline{R}_{ij}^{2}$$

RV=0.003

The Aug-rank coefficient, r_a,

a measure of

- the closeness of pairs to the RTPA
- agreement in aug-ranks ➔ homogeneity

$$r_a = 1 - \frac{n^3 RV}{(n^3 - n) - n^3 IV}$$

Adjusted for ties (internal variance) IV

Conclusions PA: 77%

r_a: 0.997 negligible individual disagreement

Homogeneous group

RTPA reveals slight systematic disagreement,

RP: -0.013, RC: 0.064

y \ X	None	Negli-gible	Mode-rate	Rather severe	Very severe
Very severe				3	3
Rather severe			5	36	7
Mode-rate		2	33	5	
Negli-gible		5			
None	1	1			

r_a

Member of correlation family
Correlation of aug-ranks

Svensson r_a:

$$r_a = \frac{(n^3 - n) - n^3 IV - 6\sum_{i=1}^{m}\sum_{j=1}^{m}(x_{ij} \cdot \Delta\bar{R}_{ij}^2)}{\sqrt{(n^3 - n)^2 - 2n^3(n^3 - n)IV + \left(n^3 \cdot IV\right)^2}}$$

Spearman r_s

$$r_s = \frac{n^3 - n - 6\sum_{v=1}^{m} d_v^2 - \tfrac{1}{2}(t^{(X)} + t^{(Y)})}{\sqrt{\left(n^3 - n\right)^2 - (t^{(X)} + t^{(Y)})(n^3 - n) + t^{(X)}t^{(Y)}}}$$

Different applications, interpretations

Association IS NOT Agreement

r and r_s

are sometimes used

as reliability measures, but

This is WRONG

The aug-rank methods are

designed for

separate analysis of the

individual and systematic

components of (dis)agreement

Y \ X	A	B	C	D	total
D				3	3
C				31	31
B			12		12
A	2	11			13
total	2	11	12	34	59

PA, 8%
The RTPA:
r_a, 1

Systematic d-a
RP, -0.53; RC, 0.36

r_s, 0.97 κ, -0.12

Y \ X	A	B	C	D	tot
D		1	3	30	34
C	1	1	7	3	12
B	1	6	1	1	9
A	2	1	1		4
	4	9	12	34	59

PA, 76%
Individual d-a:
r_a, 0.78

Marginal homogeneity
RP=RC=0

r_s, 0.77; κ, 0.60

Y \ X	A	B	C	D	tot
D			2		2
C		1		27	28
B	1	1	8	5	15
A	1	9	3	1	14
	2	11	13	33	59

PA, 3%
Individual d-a:
r_a, 0.90

Systematic d-a:
RP, -0.56; RC, 0.35

r_s, 0.73; κ, -0.18

Some references

- www.oru.se/esi/svensson

Svensson E.

•A coefficient of agreement adjusted for bias in paired ordered categorical data. Biometrical Journal 1997;39:643-57

•Application of a rank-invariant method to evaluate reliability of ordered categorical assessments. Journal of Epidemiology and Biostatistics, 1998; 3 (4):403-9

•Ordinal invariant measures for individual and group changes in ordered categorical data. Statistics in Medicine 1998;17(24): 2923-36

•Comparison of the quality of assessments using continuous and discrete ordinal rating scales. Biometrical Journal 2000;42:417-34

•Concordance between ratings using different scales for the same variable. Statistics in Medicine 2000;19(24):3483-96

•Construction of a single global scale for multi-item assessments of the same variable. Statistics in Medicine 2001;20:3831-46

•Guidelines to statistical evaluation of data from ratings scales and questionnaires. Journal of Rehabilitation Medicine 2001;33: 47-8