L'importanza di essere significante: Un esempio basato sul test della Torre di Londra

Ottavia M. Epifania, Luca Stefanutti, Pasquale Anselmi, Andrea Brancaccio, Debora de Chiusole

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata,
Università di Padova

La psicometria tra oggi e domani:
Sfide e nuovi orizzonti

20 Giugno 2024

The ratio between the measures of a and b is constant and independent of the measurement unit:

$$\frac{\varphi(a)}{\varphi(b)} = \frac{\varphi'(a)}{\varphi'(b)},$$

where φ and φ' are two different scales of measurement of the same variable [1].

[1] Strictly referring to extensive physical measures

The ratio between the measures of a and b is constant and independent of the measurement unit:

$$\frac{\varphi(a)}{\varphi(b)} = \frac{\varphi'(a)}{\varphi'(b)},$$

where φ and φ' are two different scales of measurement of the same variable [1].

Meaningful comparisons

The comparison between a and b is meaningful if it is invariant under all the unit transformations.

[1] Strictly referring to extensive physical measures

The ratio between the measures of a and b is constant and independent of the measurement unit:

$$\frac{\varphi(a)}{\varphi(b)} = \frac{\varphi'(a)}{\varphi'(b)},$$

where φ and φ' are two different scales of measurement of the same variable [1].

Meaningful comparisons

The comparison between a and b is meaningful if it is invariant under all the unit transformations.

Meaningful comparisons 2.0

Given that there is a difference between a and b, is this difference significant (or not) regardless of the scales of measurement?

[1] Strictly referring to extensive physical measures

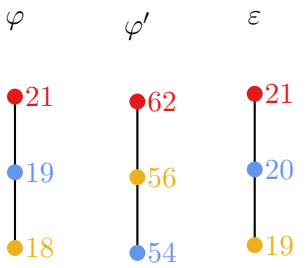
$\varphi(P) = [0, 1, 2, 3]$ $\varphi'(P) = [0, 2, 4, 10]$ $\varepsilon(P) = [0, 2, 2, 3]$

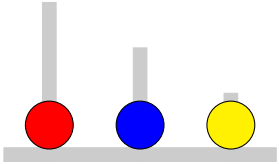
$\varphi(P) = [0, 1, 2, 3]$ $\varphi'(P) = [0, 2, 4, 10]$ $\varepsilon(P) = [0, 2, 2, 3]$

	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9
				φ					
Joe	0	1	2	2	2	3	3	3	3
Jane	0	2	2	2	3	3	3	3	3
Max	0	1	0	2	3	3	3	3	3
				φ'					
Joe	0	2	4	4	4	10	10	10	10
Jane	0	4	4	4	10	10	10	10	10
Max	0	2	0	4	10	10	10	10	10
				ϵ					
Joe	0	2	2	2	2	3	3	3	3
Jane	0	2	2	2	3	3	3	3	3
Max	0	2	0	2	3	3	3	3	3

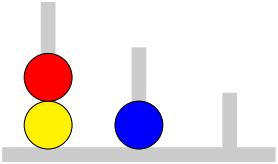
$\varphi(P) = [0, 1, 2, 3]$ $\varphi'(P) = [0, 2, 4, 10]$ $\varepsilon(P) = [0, 2, 2, 3]$

	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9
				φ					
Joe	0	1	2	2	2	3	3	3	3
Jane	0	2	2	2	3	3	3	3	3
Max	0	1	0	2	3	3	3	3	3
				φ'					
Joe	0	2	4	4	4	10	10	10	10
Jane	0	4	4	4	10	10	10	10	10
Max	0	2	0	4	10	10	10	10	10
				ϵ					
Joe	0	2	2	2	2	3	3	3	3
Jane	0	2	2	2	3	3	3	3	3
Max	0	2	0	2	3	3	3	3	3



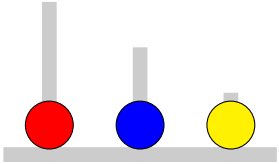


Starting configuration

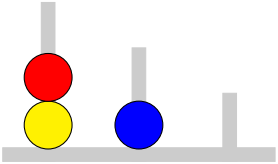


Goal configuration

The Tower of London Test (ToL Test)



Starting configuration



Goal configuration

Problem	Minimum moves	Alternative paths
Example	2	1
1	2	1
2	2	1
3	3	2
4	3	1
5	4	2
6	4	1
7	4	1
8	4	1
9	5	2
10	5	1
11	5	1
12	5	2

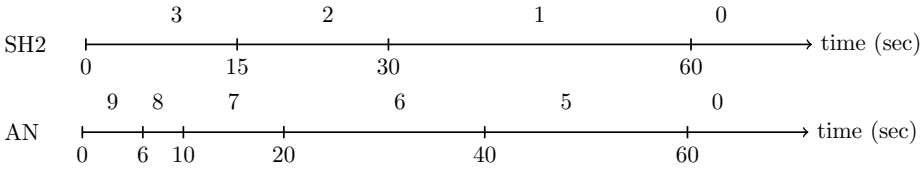
Scoring system	First attempt	Second attempt	Third attempt	Fourth on	Total sum score
KR	3	2	1	0	0 – 36
SH1	1	0			0 – 12

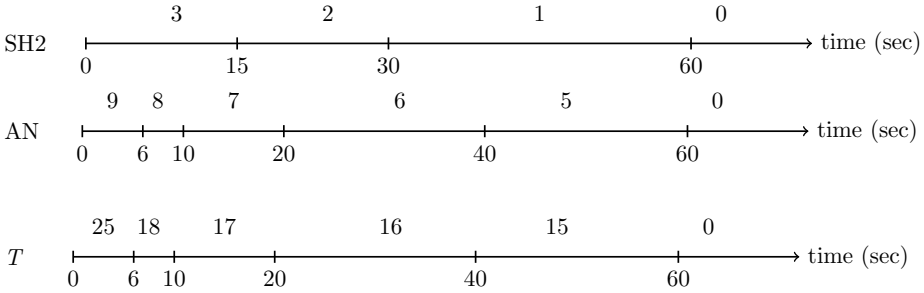
Scoring system	First attempt	Second attempt	Third attempt	Fourth on	Total sum score
KR	3	2	1	0	0 – 36
SH1	1	0			0 – 12

Scoring system	First attempt	Second attempt	Third attempt	Fourth on	Total sum score
P1	3	2		0	0 – 36
P1′	3	1		0	0 – 36

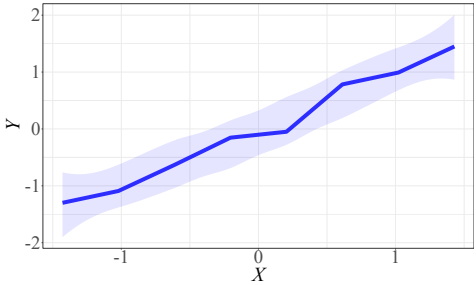
Scoring system	First attempt	Second attempt	Third attempt	Fourth on	Total sum score
KR	3	2	1	0	0 – 36
SH1	1	0			0 – 12

Scoring system	First attempt	Second attempt	Third attempt	Fourth on	Total sum score
P1	3	**2**		0	0 – 36
P1′	3	**1**		0	0 – 36

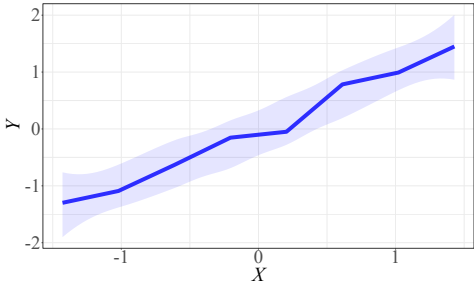

3
2
1
0
SH2
time (sec)
0
15
30
60
9
8
7
6
5
0
AN
time (sec)
0
6
10
20
40
60


3
2
1
0
SH2
time (sec)
0
15
30
60
9
8
7
6
5
0
AN
time (sec)
0
6
10
20
40
60
25
18
17
16
15
0
T
time (sec)
0
6
10
20
40
60

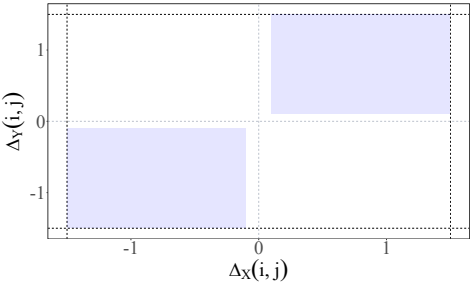
Monotonic relation



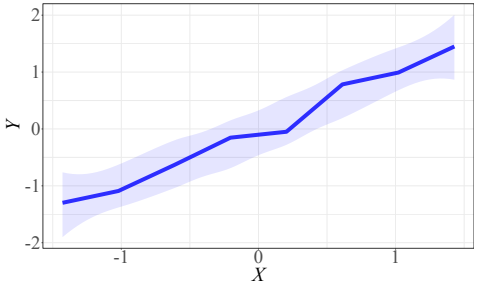
Monotonic relation



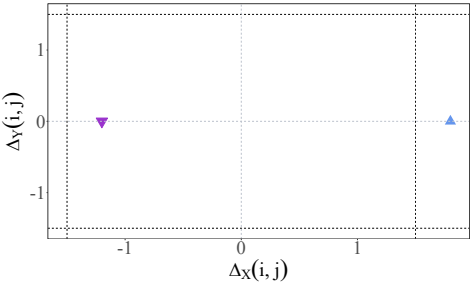
Distances and inversions



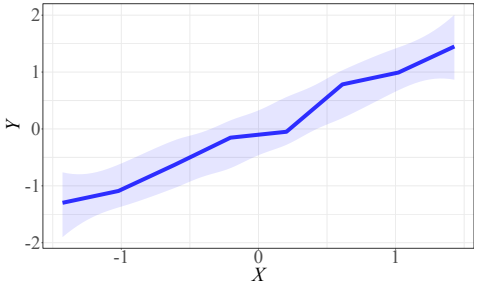
Monotonic relation



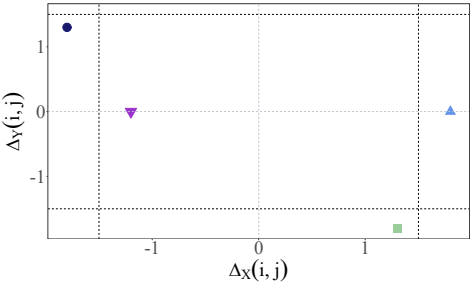
Distances and inversions



Monotonic relation

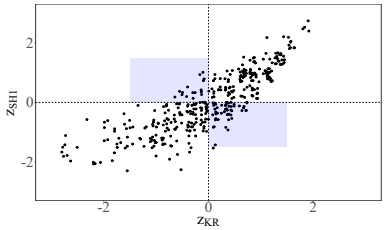


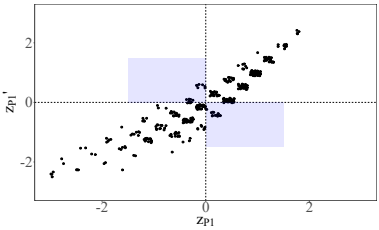
Distances and inversions



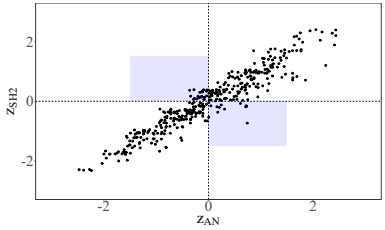
Results: Monotonic relation

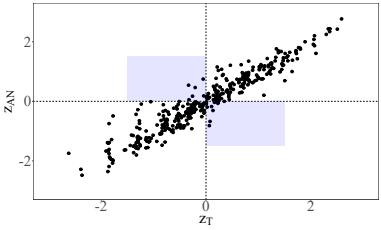
Attempt-based SMs



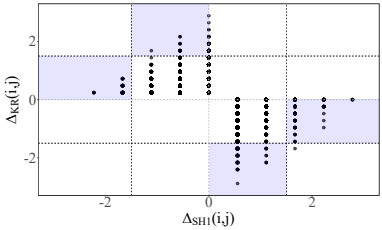


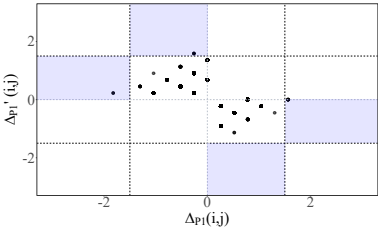
Latency-based SMs



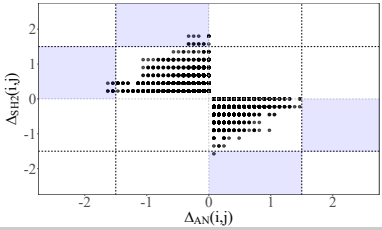


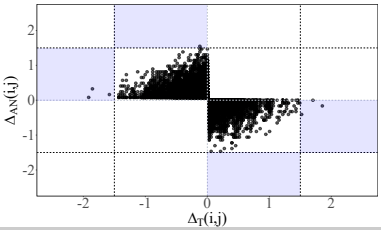
Attempt-based SMs





Latency-based SMs





H_0: $\mu_{g1} - \mu_{g2} = 0$

H_1: $\mu_{g1} - \mu_{g2} \neq 0$

t-test on the standardized scores considering different grouping variables:

Grouping variable	n_1	n_2
Gender	199	196
Administration order	202	193
Administration modality	211	184
Schooling years	171	224

	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

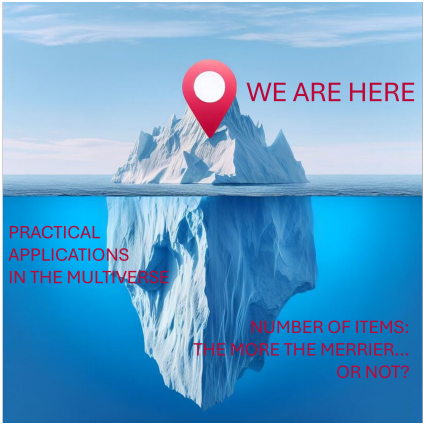
	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

	KR	SH1	P1	P1'
	d	*d*	*d*	*d*
Gender	1.84	2.11*	1.69	2.03*
	0.19	0.21	0.17	0.20
Test order	−0.15	0.80	−0.48	0.28
	−0.01	0.08	−0.05	0.03
Adm. Modality	−2.85**	−1.93	−2.69**	−2.35*
	−0.29	−0.19	−0.27	−0.24
Schooling	3.95***	3.56***	3.82***	3.85***
	0.39	0.36	0.38	0.39

	SH2	AN	T
	d	*d*	*d*
Gender	1.64	1.88	2.10*
	0.17	0.19	0.21
Test order	0.37	0.99	0.95
	0.04	0.10	0.10
Adm. Order	−2.90**	−2.33*	−2.84**
	−0.29	−0.23	−0.29
Schooling	5.52***	5.32***	5.13***
	0.56	0.54	0.52

	SH2	AN	T
	d	*d*	*d*
Gender	1.64	1.88	2.10*
	0.17	0.19	0.21
Test order	0.37	0.99	0.95
	0.04	0.10	0.10
Adm. Order	−2.90**	−2.33*	−2.84**
	−0.29	−0.23	−0.29
Schooling	5.52***	5.32***	5.13***
	0.56	0.54	0.52

	SH2	AN	T
	d	*d*	*d*
Gender	1.64	1.88	2.10*
	0.17	0.19	0.21
Test order	0.37	0.99	0.95
	0.04	0.10	0.10
Adm. Order	−2.90**	−2.33*	−2.84**
	−0.29	−0.23	−0.29
Schooling	5.52***	5.32***	5.13***
	0.56	0.54	0.52


WE ARE HERE
PRACTICAL
APPLICATIONS
IN THE MULTIVERSE
NUMBER OF ITEMS:
THE MORE THE MERRIER...
OR NOT?

Are we sure sum scores are a good idea...?

PSYCHOMETRIKA—VOL. 89, NO. 1, 84–117
MARCH 2024
https://doi.org/10.1007/s11336-024-09964-7

RECOGNIZE THE VALUE OF THE SUM SCORE, PSYCHOMETRICS' GREATEST ACCOMPLISHMENT

KLAAS SIJTSMA
TILBURG UNIVERSITY

JULES L. ELLIS
OPEN UNIVERSITY OF THE NETHERLANDS

DENNY BORSBOOM
UNIVERSITY OF AMSTERDAM

???

??

Are we sure sum scores are a good idea...?

PSYCHOMETRIKA—VOL. 89, NO. 1, 84–117
MARCH 2024
https://doi.org/10.1007/s11336-024-09964-7

RECOGNIZE THE VALUE OF THE SUM SCORE, PSYCHOMETRICS' GREATEST ACCOMPLISHMENT

KLAAS SIJTSMA
TILBURG UNIVERSITY

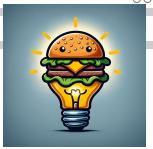
JULES L. ELLIS
OPEN UNIVERSITY OF THE NETHERLANDS

DENNY BORSBOOM
UNIVERSITY OF AMSTERDAM

???

??

Sum scores of ordinal data bring to a multiverse of contrasting results

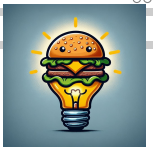


Sum scores of ordinal data bring to a multiverse of contrasting results

Increasing the number of items does not solve the issue.... it worsens it!

Meaningfulness of psychological measures and reproducibility are interlaced

Research founded by the project “Computerized, Adaptive and Personalized Assessment of Executive Functions and Fluid Intelligence” (PRIN 2020, Prot. 20209WKCLL, P.I. Prof. Luca Stefanutti)



Sum scores of ordinal data bring to a multiverse of contrasting results

Increasing the number of items does not solve the issue.... it worsens it!

Meaningfulness of psychological measures and reproducibility are interlaced

Bright side:

Sum scores of truly dichotomous data (i.e., true vs. false, correct vs. incorrect) are meaningful

Research founded by the project “Computerized, Adaptive and Personalized Assessment of Executive Functions and Fluid Intelligence” (PRIN 2020, Prot. 20209WKCLL, P.I. Prof. Luca Stefanutti)