

The European Commission's science and knowledge service

Joint Research Centre



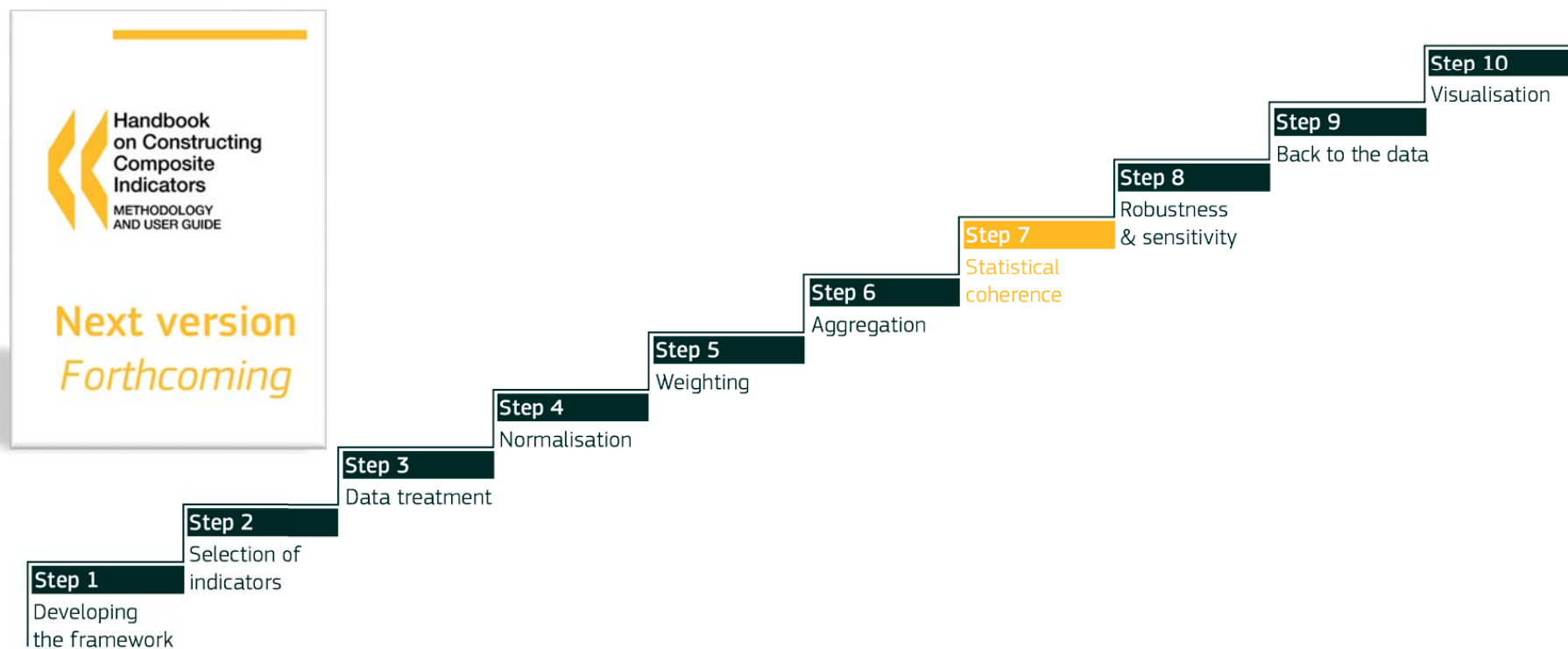
Step 7: Statistical Coherence

Simple Correlations and Cross Correlations

Pablo de Pedraza

COIN 2019 - 17th JRC Annual Training on Composite Indicators & Scoreboards
04-08/11/2019, Ispra (IT)

Ten steps



How Correlations can help?

1 indicator level	2 Component level	3 Dimension level	Index
Indicator 1(X)	Component 1(Y)	Dimension 1	Index
Indicator 2			
Indicator 3			
Indicator 4	Component 2		
Indicator 5			
Indicator 6			
Indicator 7	Component 3	Dimension 2	
Indicator 8			
Indicator 9			
Indicator 10	Component 4		
Indicator 11			

CORRELATION

$$\rho(x,y) = \text{Cov}(x,y) / (\sigma_x \sigma_y)$$

To which extend two variables are **linearly** related

1 perfect positive linear correlation.

-1 perfect negative linear correlation.

0 no linear correlation.

ρ^2 is the fraction of the variance in Y that is explained by X in a simple linear regression.

Statistically **significant** (5%) if $> 2/\sqrt{N}$

How Correlations can help?

1 indicator level	2 Component level	3 Dimension level	Index
Indicator 1	Component 1	Dimension 1	Index
Indicator 2			
Indicator 3			
Indicator 4	Component 2		
Indicator 5			
Indicator 6			
Indicator 7	Component 3	Dimension 2	
Indicator 8			
Indicator 9			
Indicator 10	Component 4		
Indicator 11			



COIN Tips

7

Check whether indicators:

- Are not related or negatively related to indicators in the same dimension < -0.3
- Are under-represented in their dimension: < 0.3
- Are negatively related to the composite indicator: < -0.3
- Does any indicators dominate the framework: > 0.95

What responses we can find?

- Are indicators allocated in the **right dimension**?
- Are some indicators **over or under represented** in the aggregate **Index**?
- Up to what **level** should we aggregate?

Social Progress Index (SPI) example (Norlén & Caperna 2018 JRC Audit)

Indicator level	Component level	Dimension level	Index
Indicators 1-5	Component 1 Nutrition & basic medical health	Dimension 1 Basic human needs	Social Progress Index
Indicators 6-9	Component 2 Water & Sanitation		
Indicators 10-12	Component 3 Shelter		
Indicators 13-16	Component 4 Personal Safety		
Indicators 17-21	Component 5 Access to basic knowledge	Dimension 2 Foundations of wellbeing	
Indicators 22-25	Component 6 Access to Information and Communication		
Indicators 26-29	Component 7 Health and Wellness		
Indicators 30-33	Component 8 Environmental Quality		
Indicators 34-38	Component 9 Personal Rights	Dimension 3 Opportunity	
Indicators 39-42	Component 10 Personal Freedom and Choice		
Indicators 43-47	Component 11 Inclusiveness		
Indicators 48-51	Component 12 Access to advanced education		

SPI example (Norlén & Caperna 2018 JRC Audit)

Indicator level
Indicators 1-5
Indicators 6-9
Indicators 10-12
Indicators 13-16
Indicators 17-21
Indicators 22-25
Indicators 26-29
Indicators 30-33
Indicators 34-38
Indicators 39-42
Indicators 43-47
Indicators 48-51

What responses we can find?
-Are indicators allocated in the **right dimension**?

A **Big table** of correlations of 51 SPI indicators. **Conclusions:** Strong & high correlations.

Indicators within any component strong **correlations between them** (majority over 0,60).

Exception: Percentage of biomes (naturally occurring community of flora and fauna) in protected areas (ind. 33) in the **environmental quality component (C8)** < 0,30

Concerns:
Highest correlations in Foundations of Wellbeing (D2) but, **not allocated in the same component (C):**
-Indicators measuring education equality (ind. 21 in C5 access to knowledge) and health equality (ind. 29 in C7 health) = 0,92

-Indicators of internet users (ind. 23) and access to essential health (ind. 28) = 0,91

SPI example (Norlén & Caperna 2018 JRC Audit)

What responses we can find?

7

-Are indicators allocated in the right dimension?

-Are some indicators over or under represented in the aggregate score?

What responses we can find?

7

-Are indicators allocated in the right dimension?

-Are some indicators over or under represented in the aggregate score?

Indicator			C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	D1	D2	D3	Index	
D1: Basic Human Needs	C2: Water and Sanitation	Access to basic drinking water	ind.01	0,85	0,75	0,77	0,54	0,67	0,73	0,69	0,59	0,43	0,64	0,44	0,67	0,80	0,72	0,64	0,76
		Water supply rate	ind.02	0,94	0,88	0,88	0,54	0,88	0,80	0,76	0,77	0,40	0,80	0,39	0,72	0,90	0,86	0,68	0,85
		Water quality	ind.03	0,93	0,87	0,88	0,58	0,90	0,82	0,77	0,78	0,42	0,83	0,43	0,77	0,91	0,88	0,72	0,87
		Water pollution	ind.04	0,88	0,83	0,84	0,65	0,82	0,82	0,82	0,76	0,52	0,81	0,58	0,80	0,88	0,87	0,79	0,88
	C3: Shelter	Communicable diseases	ind.05	0,90	0,79	0,82	0,51	0,75	0,72	0,69	0,64	0,32	0,66	0,28	0,63	0,84	0,75	0,56	0,75
		Access to basic drinking water	ind.06	0,88	0,93	0,89	0,55	0,83	0,83	0,75	0,73	0,46	0,76	0,45	0,72	0,91	0,85	0,70	0,86
		Access to piped water	ind.07	0,85	0,89	0,87	0,54	0,77	0,76	0,79	0,75	0,43	0,78	0,49	0,71	0,88	0,82	0,71	0,84
		Access to basic sanitation facilities	ind.08	0,91	0,96	0,92	0,51	0,86	0,82	0,78	0,73	0,33	0,83	0,36	0,78	0,93	0,86	0,68	0,86
	C4: Personal Safety	Rural open defecation	ind.09	0,68	0,85	0,69	0,37	0,72	0,67	0,58	0,59	0,25	0,64	0,23	0,62	0,74	0,69	0,51	0,68
		Access to electricity	ind.10	0,90	0,89	0,93	0,46	0,82	0,79	0,71	0,70	0,33	0,74	0,34	0,70	0,89	0,81	0,62	0,81
		Quality of electricity supply	ind.11	0,80	0,81	0,91	0,66	0,81	0,81	0,85	0,78	0,41	0,83	0,54	0,80	0,89	0,87	0,76	0,88
		Household air pollution deaths	ind.12	0,89	0,88	0,95	0,49	0,82	0,78	0,83	0,78	0,35	0,80	0,38	0,74	0,90	0,86	0,67	0,85
	C5: Health	Political killings and torture	ind.13	0,54	0,48	0,48	0,78	0,51	0,64	0,58	0,57	0,85	0,54	0,77	0,46	0,61	0,62	0,77	0,69
		Perceived criminality	ind.14	0,46	0,43	0,46	0,82	0,51	0,49	0,52	0,41	0,43	0,56	0,57	0,47	0,58	0,52	0,58	0,58
		Traffic deaths	ind.15	0,53	0,43	0,47	0,73	0,50	0,51	0,63	0,58	0,61	0,44	0,54	0,52	0,57	0,59	0,62	0,62
		Homicide rate	ind.16	0,40	0,32	0,37	0,74	0,45	0,44	0,47	0,37	0,25	0,41	0,34	0,49	0,48	0,46	0,44	0,48

SPI example (Norlén & Caperna 2018 JRC Audit)

What responses we can find?

- Are indicators allocated in the right dimension?
- Are some indicators over or under represented in the aggregate score?

		Indicator		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	D1	D2	D3	Index
D2: Foundations of Wellbeing	C6: Access to Info & Commun	Mobile telephone subscriptions	ind.17	0,84	0,82	0,83	0,46	0,91	0,76	0,70	0,75	0,36	0,81	0,39	0,75	0,83	0,84	0,68	0,82
		Internet users	ind.18	0,65	0,67	0,61	0,46	0,80	0,66	0,54	0,60	0,38	0,60	0,37	0,59	0,66	0,69	0,57	0,67
		Access to online governance	ind.19	0,91	0,90	0,90	0,56	0,93	0,83	0,77	0,75	0,45	0,83	0,49	0,79	0,91	0,88	0,75	0,89
		Access to independent media	ind.20	0,59	0,55	0,57	0,46	0,70	0,51	0,48	0,52	0,30	0,51	0,25	0,48	0,60	0,59	0,45	0,57
	C7: Health and Wellness	Life expectancy at 60 years	ind.21	0,62	0,59	0,63	0,69	0,73	0,66	0,73	0,58	0,50	0,69	0,63	0,64	0,69	0,72	0,72	0,74
		Non-communicable disease deaths	ind.22	0,69	0,70	0,67	0,44	0,67	0,78	0,57	0,59	0,48	0,59	0,33	0,54	0,70	0,71	0,57	0,69
		Access to essential health services	ind.23	0,87	0,87	0,88	0,66	0,84	0,92	0,86	0,80	0,52	0,88	0,56	0,88	0,91	0,92	0,83	0,93
		Access to quality healthcare	ind.24	0,76	0,72	0,76	0,56	0,75	0,88	0,74	0,70	0,49	0,75	0,48	0,81	0,78	0,83	0,75	0,82
	C8: Environmental Quality	Outdoor air pollution deaths	ind.25	0,64	0,62	0,61	0,63	0,63	0,85	0,70	0,68	0,74	0,68	0,68	0,66	0,68	0,78	0,81	0,79
		Wastewater treatment	ind.26	0,81	0,78	0,83	0,63	0,73	0,81	0,96	0,82	0,50	0,75	0,55	0,77	0,84	0,89	0,75	0,87
		Greenhouse gas emissions	ind.27	0,54	0,53	0,62	0,47	0,49	0,58	0,85	0,66	0,36	0,50	0,43	0,53	0,60	0,69	0,53	0,63
		Biome protection	ind.28	0,89	0,85	0,88	0,73	0,86	0,89	0,96	0,83	0,51	0,87	0,59	0,87	0,92	0,95	0,83	0,94
			ind.29	0,72	0,69	0,73	0,73	0,75	0,74	0,83	0,64	0,52	0,76	0,66	0,71	0,79	0,79	0,77	0,82
			ind.30	0,73	0,70	0,73	0,59	0,74	0,71	0,79	0,88	0,58	0,74	0,60	0,70	0,76	0,83	0,77	0,82
			ind.31	0,78	0,79	0,81	0,52	0,77	0,82	0,75	0,83	0,44	0,77	0,47	0,81	0,81	0,85	0,73	0,83
			ind.32	0,57	0,54	0,59	0,38	0,54	0,55	0,56	0,77	0,32	0,58	0,32	0,47	0,58	0,64	0,50	0,60
			ind.33				0,32		0,28	0,30	0,48	0,40		0,40	0,24		0,34	0,36	0,31

SPI example (Norlén & Caperna 2018 JRC Audit)

7

What responses we can find?

- Are indicators allocated in the right dimension?
- Are some indicators over or under represented in the aggregate score?

Summary:
All indicators are positively and strongly correlated with their components.

About 75% (37/51) of the correlations are greater than 0,80

About 25% (12/51) are very strong with correlations over 0,92.

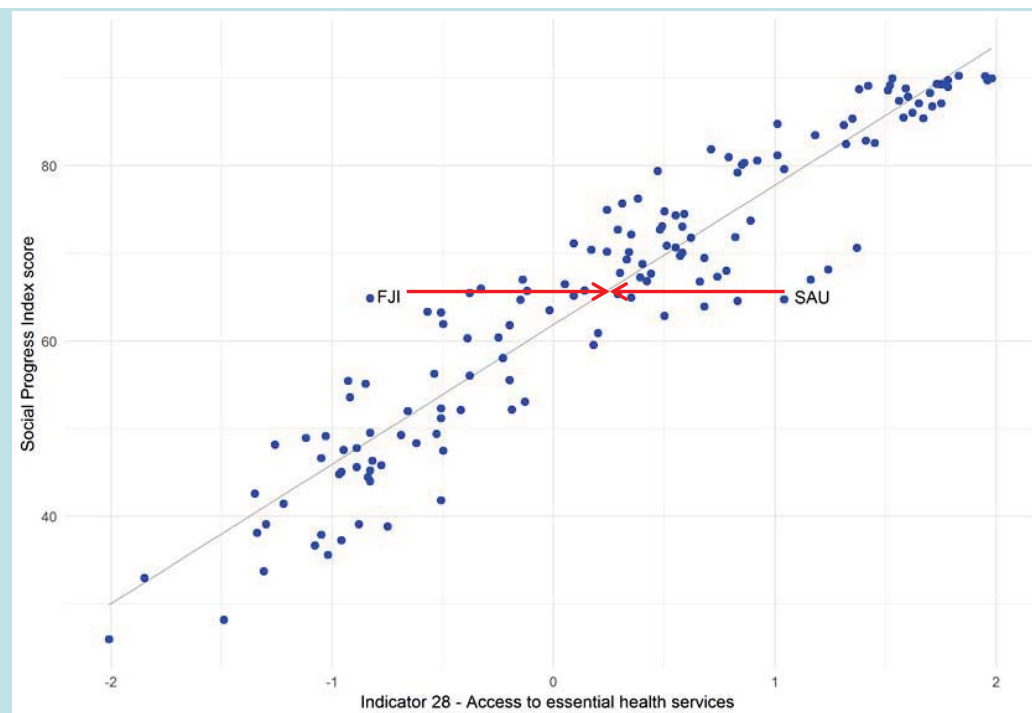
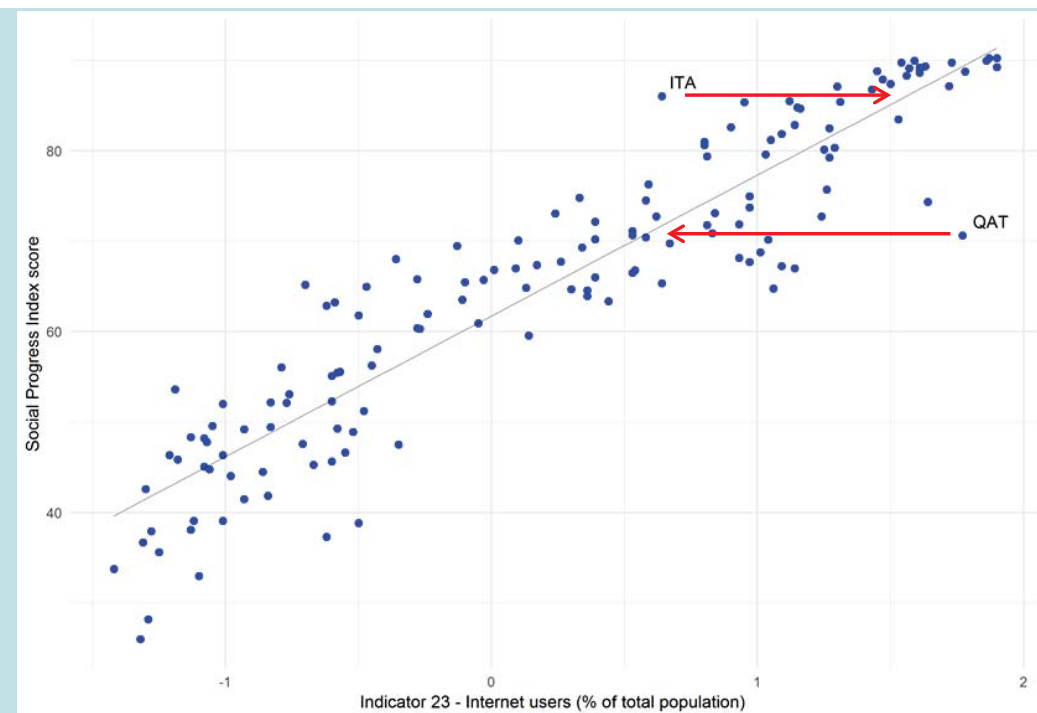
These strong correlations are found in all three dimensions but most cases are found in the first two.

		Indicator	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	D1	D2	D3	Index
D3: Opportunity	Personal Freedom and Choice	Early Marriage																0,67
		Satisfied demand for corruption																0,51
		Corruption																0,36
		Acceptance of gays and lesbians	ind.43	0,51	0,49	0,53	0,52	0,50	0,61	0,63	0,62	0,57	0,64	0,77	0,63	0,56	0,64	0,76
																		0,67
	Education	number of globally ranked universities	ind.50	0,59	0,58	0,62	0,43	0,57	0,66	0,63	0,59	0,31	0,66	0,41	0,84	0,62	0,66	0,65
		% tertiary stud. globally ranked universities	ind.51	0,58	0,58	0,61	0,54	0,57	0,68	0,68	0,65	0,44	0,69	0,54	0,80	0,64	0,69	0,73
																		0,67
																		0,71

All the indicators are correlating positively and significantly with the overall index (many > (0,7)² = 50% of the total variation is captured by the overall index, SPI).

The universal health coverage index indicator (ind. 28) and the indicator estimating the number of internet users (ind. 23) in the second dimension Foundations of Wellbeing are correlating the highest with the SPI index (correlations 0,94 and 0,93, respectively).

SPI example (Norlén & Caperna 2018 JRC Audit)



SPI example (Norlén & Caperna 2018 JRC Audit)

What responses we can find?

7

-Are components allocated in the right dimension?

and index	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	D1	D2	D3	Index
Nutrition and Basic Medical Care C1	1,00												0,96	0,91	0,75	0,91
Water and Sanitation C2	0,91	1,00											0,95	0,89	0,72	0,89
Shelter C3	0,93	0,93	1,00										0,96	0,91	0,73	0,91
Personal Safety C4	0,63	0,54	0,58	1,00									0,73	0,71	0,78	0,77
Access to Basic Knowledge C5	0,89	0,87	0,88	0,64	1,00								0,91	0,92	0,78	0,91
Access to Information and Communications C6	0,86	0,85	0,85	0,68	0,85	1,00							0,90	0,95	0,87	0,94
Health and Wellness C7	0,83	0,80	0,86	0,71	0,79	0,84	1,00						0,88	0,93	0,81	0,91
Environmental Quality C8	0,79	0,77	0,81	0,62	0,79	0,81	0,83	1,00					0,83	0,92	0,81	0,89
Personal Rights C9	0,46	0,41	0,39	0,69	0,48	0,65	0,52	0,58	1,00				0,52	0,61	0,83	0,67
Personal Freedom and Choice C10	0,83	0,83	0,85	0,64	0,85	0,85	0,81	0,80	0,51	1,00			0,88	0,89	0,87	0,91
Inclusiveness C11	0,47	0,42	0,45	0,73	0,52	0,61	0,62	0,60	0,81	0,62	1,00		0,55	0,63	0,87	0,71
Access to Advanced Education C12	0,79	0,78	0,80	0,63	0,81	0,85	0,81	0,78	0,49	0,85	0,58	1,00	0,83	0,87	0,86	0,89
Basic Human Needs D1													1,00			0,96
Foundations of Wellbeing D2													0,95	1,00		0,98
Opportunity D3													0,81	0,88	1,00	0,93

C4 better with
C5, C6, ...

Potential redundancy

To be continued in PCA...

Sustainable development Goals Index (SDG) example (Papadimitriou et al 2019 JRC audit)

85 indicators	17 SDGs	Index
2 Indicators	SDG 1: No poverty	<div>SDG Index</div> <div>Economic</div> <div>Social</div> <div>Environmental</div>
7 Indicators	SDG 2: Zero hunger	
14 Indicators	SDG 3: Good Health and well Being	
3 Indicators	SDG 4: Quality Education	
4 Indicators	SDG 5: Gender Equality	
5 Indicators	SDG 6: Clean water and sanitation	
3 Indicators	SDG 7: Affordable and Clean energy	
5 Indicators	SDG 8: Decent Work and Economic Growth	
6 Indicators	SDG 9: Industry Innovation and Infrastructure	
1 Indicator	SDG 10: Reduced Inequality	
3 Indicators	SDG 11: Sustainable Cities and Communities	
6 Indicators	SDG 12: Responsible Consumption and Production	
4 Indicators	SDG 13: Climate Action	
4 Indicators	SDG 14: Life Below Water	
5 Indicators	SDG 15: Life and Land	
9 Indicators	SDG 16: Peace and Justice Strong Institutions	
4 Indicators	SDG 17: Partnership to Achieve the Goal	

Sustainable Development Goals Index (SDG) example (Papadimitriou et al 2019 JRC audit)

SDG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Index
1	1.00																	
2	0.49	1.00																
3	0.84	0.64	1.00															
4	0.77	0.61	0.84	1.00														
5	0.35	0.54	0.59	0.61	1.00													
6	0.69	0.66	0.81	0.73	0.68	1.00												
7	0.88	0.50	0.85	0.81	0.46	0.71	1.00											
8	0.50	0.60	0.68	0.62	0.59	0.63	0.51	1.00										
9	0.66	0.66	0.82	0.67	0.59	0.75	0.68	0.63	1.00									
10	0.36	0.29	0.36	0.17	0.01	0.21	0.19	0.24	0.38	1.00								
11	0.54	0.46	0.68	0.66	0.64	0.69	0.59	0.53	0.55	0.10	1.00							
12	-0.59	-0.53	-0.76	-0.59	-0.52	-0.67	-0.60	-0.46	-0.86	-0.32	-0.50	1.00						
13	-0.28	-0.17	-0.33	-0.32	-0.18	-0.19	-0.27	-0.10	-0.32	-0.05	-0.13	0.50	1.00					
14	-0.17	-0.01	-0.11	-0.10	0.09	-0.04	-0.13	0.06	-0.05	-0.19	-0.04	0.04	0.02	1.00				
15	-0.09	0.15	-0.02	-0.03	0.17	0.12	-0.05	0.14	0.09	0.01	-0.01	-0.02	0.22	0.24	1.00			
16	0.64	0.59	0.81	0.67	0.50	0.65	0.60	0.58	0.78	0.43	0.63	-0.72	-0.33	-0.11	0.04	1.00		
17	0.19	0.07	0.21	0.22	0.14	0.12	0.29	-0.03	0.11	0.01	0.16	-0.12	-0.35	-0.02	-0.08	0.14	1.00	
Index	0.84	0.71	0.93	0.86	0.67	0.86	0.86	0.73	0.83	0.40	0.73	-0.68	-0.20	-0.01	0.14	0.79	0.24	1.00

>0,7 the index captures at least 50% of the variation in the underlying goal and vice versa (SDG 1 to SDG 9, SDG 11)

Negative relationship with the index
SDG 12: Responsible Consumption and Production
SDG 13: Climate Action

No correlation

Very high may suggest that SDG 3 is driving the index or telling the entire story

What's wrong with negative correlations?

$X1$ & $X2$
negatively correlated

Normalization
(Distance to the best performer):
 $I1 = (X1/\max X1) \cdot 100$ & $I2 = (X2/\max X2) \cdot 100$

$$Y = .5 \cdot I_1 + .5 \cdot I_2$$

Country	$X1$	$X2$	$I1$	$I2$	Y	Rank
H	2 000	500	100	100	100	1
A	160	435	8	87	47.5	2
B	400	370	20	74	47.0	3
C	640	305	32	61	46.5	4
D	880	240	44	48	46.0	5
E	1 120	175	56	35	45.5	6
F	1 360	110	68	22	45.0	7
G	1 600	45	80	9	44.5	8

2.- What's wrong with negative correlations?

$X1$ & $X2$
negatively correlated

Normalization
(Distance to the best performer):
 $I1 = (X1/\max X1) \cdot 100$ & $I2 = (X2/\max X2) \cdot 100$

$$Y = .5 \cdot I_1 + .5 \cdot I_2$$

Country	$X1$	$X2$	$I1$	$I2$	Y	Rank
H	2 000	500	100	100	100	1
A	160	435	8	87	47.5	2
B	400	370	20	74	47.0	3
C	640	305	32	61	46.5	4
D	880	240	44	48	46.0	5
E	1 120	175	56	35	45.5	6
F	1 360	110	68	22	45.0	7
G	1 600	45	80	9	44.5	8

2.- What's wrong with negative correlations?

The best country improves and we recalculate everything

Country	X1	X2	I1	I2	Y	Rank
H	2 000	700	100	100	100	1
A	160	435	8	62.14	35.07	8
B	400	370	20	52.86	36.43	7
C	640	305	32	43.57	37.79	6
D	880	240	44	34.29	39.14	5
E	1 120	175	56	25	40.5	4
F	1 360	110	68	15.71	41.86	3
G	1 600	45	80	6.43	43.21	2

Previous Rank
1
2
3
4
5
6
7
8

Only the best performer (H) improves BUT the ranking gets completely reversed and country A is last as opposed to 2nd !

CONCLUSIONS

Correlations → Multi-dimensionality - Redundancy

