

Spearman's correlation

<http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf>

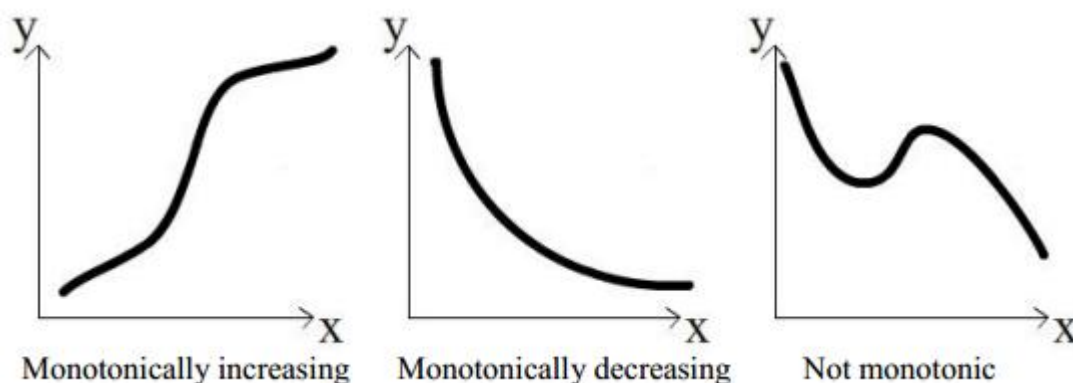
ก่อนที่จะใช้ Spearman's correlation จำเป็นต้องเข้าใจ Pearson's correlation ที่เป็นการหาความสัมพันธ์เชิงเส้นตรงของข้อมูล 2 ข้อมูล การคำนวณและทดสอบสมมุติฐานต้องมีข้อตกลงเบื้องต้นดังนี้:

- interval หรือ ratio level
- มีความสัมพันธ์เชิงเส้นตรง
- bivariate normally distributed

ถ้าไม่สามารถปฏิบัติตามข้อตกลงเบื้องต้นได้ ให้ใช้ Spearman's rank correlation

Monotonic function

เพื่อให้เข้าใจ Spearman's correlation จำเป็นต้องรู้ว่า monotonic function คืออะไร monotonic function คือ สิ่งที่ไม่เคยเพิ่มขึ้น หรือ ลดลง เมื่อตัวแปรต้นของมันเปลี่ยนแปลง กราฟต่อไปนี้แสดง monotonic function:



- monotonically increasing - เมื่อตัวแปร x เพิ่มขึ้น ตัวแปร y ไม่เคยลดลง
- monotonically decreasing - เมื่อตัวแปร x ลดลง ตัวแปร y ไม่เคยเพิ่มขึ้น
- not monotonic - เมื่อตัวแปร x เพิ่มขึ้น ตัวแปร y บางครั้งก็ลดลง บางครั้งก็เพิ่มขึ้น

Spearman's correlation coefficient

Spearman's correlation coefficient เป็นการวัดความแรงของความสัมพันธ์ monotonic ระหว่างตัวแปร 2 ตัว ใช้ตัวย่อว่า r_s ซึ่งมีค่าดังนี้ $-1 \leq r_s \leq 1$

การแปลผลคล้ายกับ Pearsons คือ เมื่อ r_s เข้าใกล้ +1 หรือ -1 เท่าไรก็มีความสัมพันธ์ monotonic มากเท่านั้น

correlation เป็น effect size และสามารถอธิบายความแรงของ correlation ได้โดยค่าของ r_s :

.00 - .19	very weak
.20 - .39	weak
.40 - .59	moderate
.60 - .79	strong
.80 - 1.0	very strong

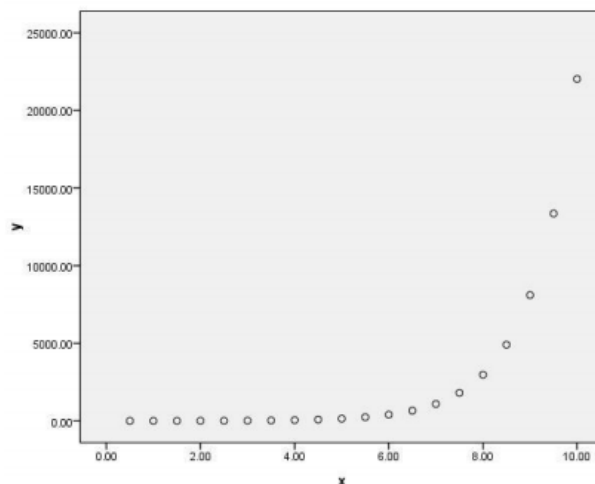
การคำนวณ Spearman's correlation coefficient และการทดสอบนัยสำคัญต้องมีข้อตกลงเบื้องต้นดังนี้:

- interval หรือ ratio level หรือ ordinal
- ความสัมพันธ์ monotonic

ไม่ต้องใช้ normality จึงเป็น nonparametric statistic

ตัวอย่าง ตารางค่า x และ y มีความสัมพันธ์ $y = \exp(x)$ จากกราฟเห็นว่ามีความสัมพันธ์แบบ increasing monotonic

Num	x	Y
1	.5	1.6
2	1.0	2.7
3	1.5	4.5
4	2.0	7.4
5	2.5	12.2
6	3.0	20.1
7	3.5	33.1
8	4.0	54.6
9	4.5	90.0
10	5.0	148.4
11	5.5	244.7
12	6.0	403.4
13	6.5	665.1
14	7.0	1096.6
15	7.5	1808.0
16	8.0	2981.0
17	8.5	4914.8
18	9.0	8103.1
19	9.5	13359.7
20	10.0	22026.5



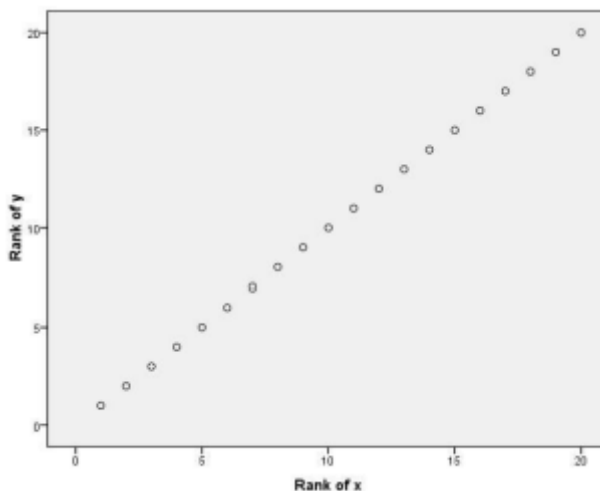
การคำนวณ Pearson's correlation ได้เป็น .699 ไม่ได้แสดงว่ามีความสัมพันธ์ของข้อมูลอย่างสมบูรณ์ แต่

Spearman's correlation ของข้อมูลนี้เป็น 1 แสดงว่ามีความสัมพันธ์ monotonic อย่างสมบูรณ์

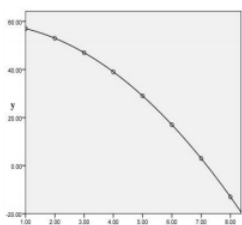
Spearman's correlation คำนวณได้โดยหา Pearson's correlation ของค่าข้อมูลเรียงลำดับจากน้อยไปมาก โดยให้ค่า 1 กับค่าที่น้อยที่สุดและ 2 กับค่าต่อไปและทำเช่นนี้ต่อไปเรื่อย ๆ

หากวาดกราฟของข้อมูลเรียงลำดับจะเห็นความสัมพันธ์เชิงเส้นตรงสมบูรณ์

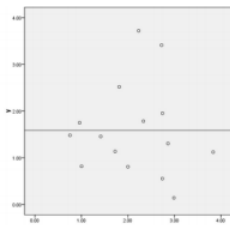
Num	X	Rank ของ x	Y	Rank ของ y
1	.5	1	1.6	1
2	1.0	2	2.7	2
3	1.5	3	4.5	3
4	2.0	4	7.4	4
5	2.5	5	12.2	5
6	3.0	6	20.1	6
7	3.5	7	33.1	7
8	4.0	8	54.6	8
9	4.5	9	90.0	9
10	5.0	10	148.4	10
11	5.5	11	244.7	11
12	6.0	12	403.4	12
13	6.5	13	665.1	13
14	7.0	14	1096.6	14
15	7.5	15	1808.0	15
16	8.0	16	2981.0	16
17	8.5	17	4914.8	17
18	9.0	18	8103.1	18
19	9.5	19	13359.7	19
20	10.0	20	22026.5	20



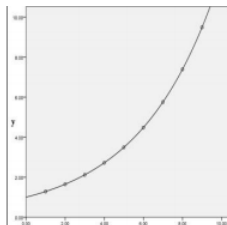
ภาพข้างล่างเป็นตัวอย่างและความสัมพันธ์แสดงค่า correlation coefficient โดย 3 รูปแรกแสดงค่า monotonic correlation แบบสุดขีด -1, 0 และ 1:



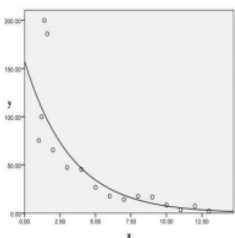
$r_s = -1$
perfect -ve
monotonic correlation



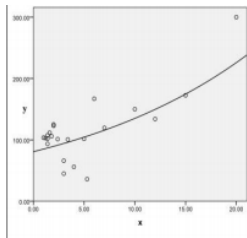
$r_s = 0$
no correlation



$r_s = 1$
perfect +ve
monotonic correlation



$r_s = -.941$
very strong -ve
monotonic correlation



$r_s = .372$
weak +ve
monotonic correlation

Spearman's rank order correlation

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

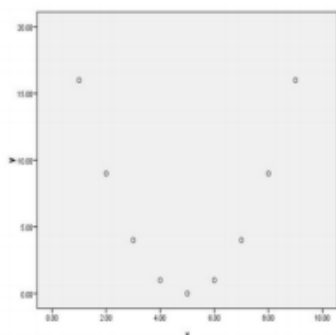
หากข้อมูลเป็น tied data

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

tied data: ข้อมูลที่มีตัวเลขซ้ำ ๆ กัน เช่น 1,2,3,4,5,5,6,7,7 มี 5 และ

7 เป็น tied data

Spearman's correlation coefficient เป็นการวัดความสัมพันธ์ monotonic ดังนั้นเมื่อค่า $r_s = 0$ ไม่ได้หมายความว่าตัวแปรไม่มีความสัมพันธ์กัน ตัวอย่างเช่น scatter plot ต่อไปนี้มี $r_s = 0$ ซึ่งหมายความว่าไม่มี (monotonic) correlation แต่มีความสัมพันธ์แบบ quadratic



$$r_s = 0$$

perfect quadratic relationship

ตัวอย่างเช่น

น้ำใต้ดินจำนวน 23 ตัวอย่าง ตรวจความเข้มข้นของยูเรเนียม (ppb) และ total dissolved solid (mg/L) ต้องการทราบว่าตัวแปรทั้งสองนี้มีความสัมพันธ์กันหรือไม่

ควรเริ่มจาก Pearson's correlation เหมาะสมในการใช้วิเคราะห์หรือจะใช้ Spearman's

	ความเข้มข้นยูเรเนียม (ppb)	TDS (mg/L)
1	678.10	.80
2	818.93	1.93
3	302.38	.97
4	1149.60	11.80
5	573.14	1.41
6	1034.55	2.41
7	633.25	3.40
8	1095.42	.98
9	1122.58	2.46
10	686.51	.26
11	1172.84	9.97
12	593.70	.37
13	1247.95	6.70
14	533.99	.09
15	605.51	1.72
16	696.96	6.76
17	1282.95	10.27
18	531.16	.13
19	788.36	2.87
20	956.06	3.10
21	1149.38	.96
22	1069.82	3.77
23	1124.17	7.09

scatter plot แสดงให้เห็นความสัมพันธ์เชิงบวกระหว่างยูเรเนียมและ TDS แต่เป็นแบบไม่ใช่เส้นตรง

Spearman's correlation ใช้ในการตัดสินความสัมพันธ์ระหว่างยูเรเนียมและ TDS พบความสัมพันธ์เชิงบวก $r_s = .71$, $n = 23$, $p < .001$

