2015 ■ vol. 11 ■ no. 1

DOI: 10.20982/tqmp.11.1.p001

Un indice d'association non-linéaire entre deux variables continues, en version non-paramétrique

A general non-linear index of association between two continuous rank-order variables

Louis Laurencelle [a]

[a] Université du Québec à Trois-Rivières

Abstract ■ Non-linear dependence between two continuous variables has been given but little consideration among statisticians to this day, and no correlation index has been contrived, apart from the semi-categorized η^2 coefficient in the anova context. Here, a non-parametric, rank-based approach is implemented, giving rise to two coefficients, R_Y, which measures the non-linear (and non-monotonic) variation of the Y series concomitant to the X series, and R_{XY}, a symmetrised measure of the non-linear correspondence between the two series. The gist of the approach resides in the postulate that, if the series are related in any manner, numerically consecutive values of one variable should be linked to values of the other variable having reduced mutual differences. R_Y and R_{XY} are presented here, with their first moments and sets of exact and approximate critical values, and they are the distribution-free counterparts of coefficients A and AS (Laurencelle, 2012) formerly presented for the normal parametric context.

Keywords ■ Non-linear dependence ; non-parametric correlation ; rank-order association index

louis.laurencelle@gmail.com

Introduction

Toutes les méthodes appliquées pour mesurer et vérifier la relation de dépendance entre deux variables statistiques reposent sur le recours à un modèle, que ce soit le modèle linéaire simple ($Y' = b{\cdot}X + a$) ou polynomial ($Y' = b_0 + b_1{\cdot}X + b_2{\cdot}X^2 + \ldots + b_k{\cdot}X^k$), la série de Fourier (en sinus et cosinus des harmoniques de la série globale), les nombreux modèles non-linéaires, tels l'exponentiel ($Y' = a{\cdot}e^{bX}$) ou le modèle en puissance ($Y' = a{\cdot}X^b$) qui sont linéarisables ou d'autres qui ne le sont pas, tels le modèle réciproque ($Y' = a{\cdot}(X + b)^{-1}$) ou les modèles plus complexes (p. ex. $Y' = a + b{\cdot}X^c$). Le chercheur aura parfois un modèle en tête, justifié par l'évidence du diagramme X :Y qu'il a devant les yeux ou par un raisonnement substantiel, sans quoi il tentera l'un ou l'autre des modèles de son répertoire habituel, en espérant démontrer en fin de compte la présence d'une relation réelle, statistiquement significative, entre ses deux variables.

Mais l'application d'une méthode basée sur un modèle repose sur deux prémisses, la première étant qu'il existe entre X et Y une relation statistique suffisante, c.-à-d. plus consistante que ne le serait un « bruit » aléatoire, et la seconde, que le modèle sied à cette relation. Or, la recherche d'un modèle et, éventuellement, la mise au point d'une méthode pour vérifier le modèle[1] supposent nécessairement qu'il existe une relation sérieuse entre X et Y. C'est pourquoi il peut être motivant d'établir cette relation – quel que soit le modèle éventuellement trouvé –, et, dans certains cas, cette preuve de relation pourra être la seule issue concluante de l'étude, faute d'un modèle identifiable.

Laurencelle (2012) s'est penché sur ce problème et, après un tour d'horizon sur certaines approches de solution parues dans la littérature (voir plus bas), il propose deux indices applicables aux cas de deux variables aléatoires continues « bien distribuées »,

[1] L'avantage du principe des moindres carrés, consistant à minimiser la somme des écarts carrés entre valeurs observées (Y_i) et valeurs prédites ($Y_i' = f(X_i)$), soit $\Sigma\ (Y_i - f(X_i))^2$, ressortit surtout aux modèles linéaires (et linéarisables) en facilitant leur solution, mais il est souvent nécessaire de concevoir un critère de minimisation approprié au problème, soit généralement $\Sigma\ |Y_i - f(X_i)|^p$, avec $p \geq 1$, de même que son algorithme de solution.

référant bien sûr à la loi normale. Les indices sont construits à partir de la « variance permutative », p^2_X, d'abord étudiée par Von Neumann et ses collaborateurs (1941a,b), soit l'estimateur de variance basé sur la différence entre les valeurs consécutives de la variable :

$$p^2_X = \frac{\sum_{i=1}^{n-1}(X_i - X_{i+1})^2}{2(n-1)}.$$

Pour démontrer la présence d'une dépendance stochastique de Y sur X, Laurencelle (2012) propose l'indice A,

$$\mathrm{A} = \frac{p^2_{Y.X}}{s^2_Y} \qquad (1)$$

pour lequel $p^2_{Y.X}$ dénote la variance permutative des *concomitantes* Y de X, c.-à-d. les valeurs de Y concomitantes aux valeurs de X réarrangées en ordre. Il propose aussi un indice symétrisé, l'indice $\mathrm{A_S}$,

$$\mathrm{A_s} = \sqrt{\frac{p^2_{X.Y}\,p^2_{Y.X}}{s^2_X\,s^2_Y}} \qquad (2)$$

qui tient lieu de « corrélation » non-linéaire entre X et Y, et qui utilise aussi $p^2_{X.Y}$, la variance permutative des concomitantes X (par rapport aux Y).

L'auteur (Laurencelle, 2012) rapporte les valeurs connues des moments pour ces deux indices de dépendance stochastique, produit des valeurs critiques prolongées par des formules d'approximation et donne deux exemples élaborés, que nous reprenons plus loin.

Le contexte non-paramétrique et les indices R_Y et R_{XY}

Les données associées à une étude particulière ne sont pas forcément distribuées normalement, auquel cas l'application des indices A et $\mathrm{A_S}$ ci-dessus est discutable, et l'absence de modèle défini peut aussi jeter un doute sur l'interprétation de la variance permutative des valeurs brutes des variables X et Y. Une approche plus prudente, - plus conservatrice aussi, faut-il dire -, consiste à faire abstraction de la loi de distribution de nos deux variables, et à les remplacer chacune par des rangs, de numéros de 1 à *n*.

Développement des indices. Soit la série bivariée { X_i , Y_i }, $i = 1$ à n, et soit sa traduction en une double série de rangs { x_i' , y_i' }, où les x_i' et y_i' occupent chacune les valeurs 1 à n. Réarrangeons la série série { x_i' , y_i' } de telle sorte que les x_i' apparaissent en ordre croissant, obtenant alors la série { i, y_i }, dans laquelle maintenant les y_i sont les rangs de Y concomitants des X (c.-à-d. y_i est la valeur de Y associée à la $i^{\text{ème}}$ plus petite valeur de X). Alors, la statistique $\mathrm{S_Y}$ est calculée comme :

$$\mathrm{S_Y} = \sum_{i=1}^{n-1} |y_i - y_{i+1}| - (n-1), \qquad (3)$$

cette quantité dénotant l'espacement (ordinal) entre les valeurs de Y associées à des valeurs consécutives de X. On peut obtenir pareillement $\mathrm{S_X}$.

Les indices $\mathrm{R_Y}$ et $\mathrm{R_{XY}}$ sont alors, simplement :

$$\mathrm{R_Y} = \frac{\mathrm{S_Y}}{\lfloor n(n-2)/2 \rfloor}\,;\ \mathrm{R_{XY}} = \frac{\mathrm{S_X} + \mathrm{S_Y}}{2 \cdot \lfloor n(n-2)/2 \rfloor};\ (4)$$

noter que, au dénominateur, l'expression $\lfloor \mathrm{U} \rfloor$ dénote la partie entière de U.

Moments et distribution des indices. L'analyse fait apparaître les caractéristiques suivantes de la statistique $\mathrm{S_Y}$ (et parallèlement $\mathrm{S_X}$) :

$$\min(\mathrm{S_X}) = 0 \qquad (5a)$$

$$\max(\mathrm{S_X}) = \lfloor n(n-2)/2 \rfloor \qquad (5b)$$

$$\mathrm{E}(\mathrm{S_X}) = (n-1)(n-2)/3 \qquad (5c)$$

$$\mathrm{var}(\mathrm{S_X}) = (n-2)(4n-7)(n+1)/90 \qquad (5d)$$

Quant à la forme distributionnelle, l'indice d'asymétrie g_1, qui vaut 0 à $n = 4$, descend jusqu'à -0,112 pour $n = 10$ et remontre lentement vers 0 selon une fonction donnée approximativement par $-0{,}474\ /\sqrt{n+7}$. À $n = 50$, il est d'à peu près –0,060, à 100, de –0,045 et à 1000, de –0,010. Valant aussi 0 à $n = 4$, l'indice d'aplatissement g_2 tombe à –0,544 à $n = 5$ et il remonte lui aussi lentement vers 0 selon la fonction approximative $-56{,}04\ /(n+7)^2 - 2{,}167/(n+7)$. La forme globale est donc approximativement *normalisée* à $n = 50$, voire avant.

La statistique $\mathrm{S_X}+\mathrm{S_Y}$ hérite, bien sûr, des caractéristiques de $\mathrm{S_X}$ présentées ci-dessus, la combinaison entraînant cependant quelque différence aux moments supérieurs, soit :

$$\min(\mathrm{S_X}+\mathrm{S_Y}) = 0 \qquad (6a)$$

$$\max(\mathrm{S_X}+\mathrm{S_Y}) = 2\times\lfloor n(n-2)/2 \rfloor \qquad (6b)$$

$$\mathrm{E}(\mathrm{S_X}+\mathrm{S_Y}) = (n-1)(n-2)\cdot 2/3 \qquad (6c)$$

$$\mathrm{var}(\mathrm{S_X}+\mathrm{S_Y}) = (n-2)(4n-7)(n+1)/45\times(1+\rho_n) \qquad (6d)$$

Tableau 1. Valeurs critiques des indices R_Y et R_{XY}. Les données ci-dessus permettent de construire et d'étudier le comportement des indices R_Y et R_{XY}. Bornés tous deux entre 0 et 1, l'espérance des deux indices tend asymptotiquement (sur n) vers ⅔ et leur variance, vers 8/(45n) pour R_Y et vers 4/(45n) pour R_{XY}. Le tableau ci-dessous présente un ensemble de valeurs critiques, aux rangs centiles 1 et 5.

	R_Y		R_{XY}	
n	0,05	0,01	0,05	0,01
4	–	–	–	–
5	0,000	–	0,071	–
6	0,167	0,083	0,208	0,125
7	0,235	0,118	0,265	0,176
8	0,292	0,167	0,313	0,208
9	0,323	0,226	0,355	0,258
10	0,325	0,250	0,375	0,275
11	0,367	0,286	0,408	0,316
12	0,383	0,300	0,417	0,333
13	0,394	0,310	0,437	0,359
14	0,405	0,333	0,446	0,369
15	0,423	0,351	0,464	0,387
16	0,429	0,357	0,469	0,402
17	0,441	0,370	0,480	0,413
18	0,451	0,382	0,486	0,424
19	0,460	0,391	0,494	0,435
20	0,467	0,400	0,500	0,442
25	0,491	0,432	0,526	0,476
30	0,512	0,457	0,543	0,498
35	0,525	0,475	0,555	0,515
40	0,536	0,488	0,564	0,527
45	0,545	0,501	0,572	0,538
50	0,552	0,510	0,578	0,546
100	0,589	0,560	0,609	0,587
250	0,620	0,601	0,633	0,619
500	0,634	0,621	0,643	0,634
750	0,640	0,629	0,648	0,640
1000	0,644	0,635	0,650	0,644
∞	2/3	2/3	2/3	2/3

En effet, le maximum et la moyenne sont doublés. Quant à la variance, en plus d'être doublée, elle profite d'une corrélation positive entre S_X et S_Y, forte (ρ_4 = 0,833) pour n = 4 et allant s'atténuant avec n croissant, selon la fonction approximative $\rho' \approx 5{,}2 / (n + 2)$. Indiquant une asymétrie négative, l'indice g_1 démarre à –0,14 pour n = 4 et descend à –0,35 pour n = 9, remontant ensuite vers 0 selon la fonction approximative $12{,}28 /(n - 4)^{3/2} - 7{,}57/(n - 4) + 0{,}23/(n - 4) - 0{,}01$; l'asymétrie, à –0,054 pour n = 100, est quasi disparue à n = 1000. Quant à g_2, après d'importantes fluctuations entre n = 4 et 9, il passe de 0,021 à n = 10 pour monter jusqu'à 0,067 à n = 14, redescendant ensuite vers 0 selon la fonction approximative $-8{,}69/(n - 4)^3 + 5{,}38/(n - 4)^2 + 0{,}08/(n - 4)$, atteignant ~ 0 vers n = 60.

Les résultats ci-dessus, aux expressions (5) et (6), sont exacts, à l'exception de l'estimateur de corrélation ρ_n. Quant aux estimations elles-mêmes, pour ρ_n, les divers indices g_1, g_2 et les valeurs critiques à venir, elles sont exactes jusqu'à n = 13, étant obtenues par énumération des 13! permutations (et contre-permutations dans le cas de la statistique S_X+S_Y), et approximées par échantillonnage Monte Carlo pour n = 14 et au-delà, au prix de plusieurs millions d'échantillons.

Il est à noter que les approximations proposées (ré-illustrées plus bas) conviennent très bien (avec une précision inférieure à 0,005) pour R_Y à partir de n = 30 et pour R_{XY} à partir de n = 50. Par ailleurs, l'interpolation recommandée selon n devrait utiliser l'argument $1 / \sqrt{n}$.

Un exemple revisité pour R_Y

Le premier exemple, tiré de Laurencelle (2012), rapporte des niveaux (fictifs) d'un paramètre, tel que température ou humidité, échantillonné à intervalle régulier pour une certaine période. Le tableau suivant présente les 50 mesures successives Y_i obtenues ; les données de temps X_i, elles, sont équidistantes (p. ex. 1, 2, 3, ..., 50) :

Y_1 à Y_5 =	-0,235	1,027	-0,805	-0,365	1,951
Y_6 à Y_{10} =	0,825	1,420	1,931	0,488	-1,540
Y_{11} à Y_{15} =	-0,724	-0,148	-1,820	-1,508	-0,530
Y_{16} à Y_{20} =	-1,439	-1,022	0,485	-1,918	1,043
Y_{21} à Y_{25} =	0,019	1,438	1,005	2,774	1,224
Y_{26} à Y_{30} =	-0,154	1,463	0,680	1,491	0,726
Y_{31} à Y_{35} =	-0,853	-1,357	-2,449	-1,009	-2,200
Y_{36} à Y_{40} =	0,468	1,085	-1,183	-0,781	0,603
Y_{41} à Y_{45} =	0,026	-0,656	-0,320	2,875	1,986
Y_{46} à Y_{50} =	1,546	0,840	1,358	1,357	1,286

La Figure 1 fait voir la variation du paramètre (Y) en fonction du temps (X).

Le graphe présenté à la figure 1 conduit à trois constatations : 1) la relation de Y à X, si elle existe, est bruyante, stochastique ; (2) il pourrait y avoir une relation cyclique, de type sinusoïdal, entre Y et X, et (3)

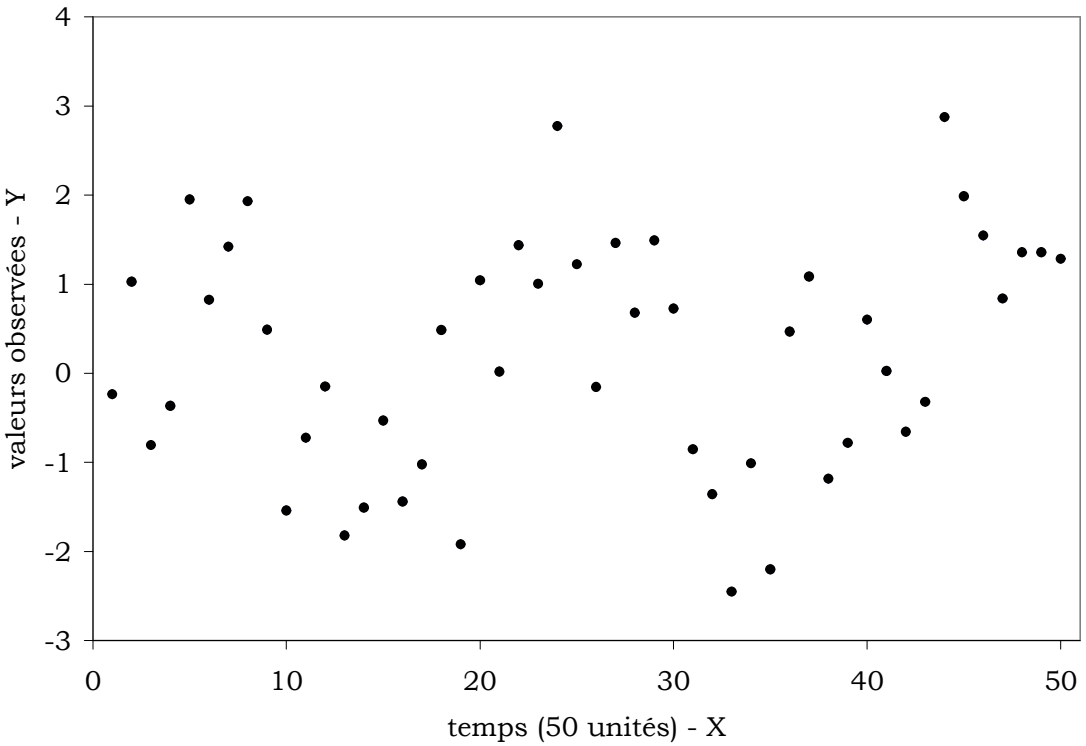

Figure 1 ■ Échantillon d'un modèle sinusoïdal de n = 50 valeurs, bruitées à 50 % (R^2 = 0,50)

on ne perçoit pas de tendance monotone évidente, ni à l'accroissement, ni à la diminution. Cette dernière constatation se confirme par le calcul du coefficient r, qui vaut ici 0,160 et n'est pas significatif au seuil bilatéral de 5 %.

Aux valeurs X_i, naturellement ordonnées dans cet exemple, correspondent des valeurs Y_i qui semblent former un tracé approximatif, une ligne de fonction, dont l'existence supposée agit en gardant les Y_i successives près l'une de l'autre. S'il n'existait nulle dépendance entre Y et X, les variations observées entre les Y_i successives seraient du même ordre que n'importe quelle des ${}_nC_2$ différences possibles parmi les n valeurs de la série. Ainsi, la dépendance de Y sur X se reflèterait par une réduction systématique des différences successives. La transformation des valeurs de la série Y en rangs, de 1 à n = 50, fournit la série inscrite dans le tableau suivant :

r_1 à r_{10} =	20	34	13	18	47	31	41	46	27	5
r_{11} à r_{20} =	15	22	4	6	17	7	10	26	3	35
r_{21} à r_{30} =	23	42	33	49	37	21	43	29	44	30
r_{31} à r_{40} =	12	8	1	11	2	25	36	9	14	28
r_{41} à r_{50} =	24	16	19	50	48	45	32	40	39	38

La sommation des intervalles successifs des rangs de la série fournit la quantité S_Y, à savoir :

$$S_Y = |20-34| + |34-13| + \ldots + |39-38| \; - \; (50-1)$$
$$= 624 - 49$$
$$= 575.$$

Pour n = 50, d'après (5c-5d), $E(S_X)$ = 784, σ^2 = 5294,6 et $\sigma \approx 72{,}454$. L'indice R_Y, soit S_X / max(S_X), vaut ici ~ 0,479, et les valeurs critiques au tableau 1 sont respectivement 0,552 (α = 0,05) et 0,510 (α = 0,01). Puisque $R_Y = 0{,}479 \leq 0{,}510$, la dépendance de Y sur X peut donc être déclarée significative au seuil de 1%, ce qui concorde avec la conclusion obtenue par Laurencelle (2012) grâce à l'indice A.

La taille n de notre exemple atteignant 50, l'approximation normale fait aussi l'affaire, selon la

formule traditionnelle :

$$z = \frac{S_Y - E(S_Y) + 0{,}5}{\sigma(S_Y)} \; ; \qquad (7)$$

la correction pour continuité doit ici être *ajoutée*, le test étant unilatéral à gauche. Le calcul donne z = (575 – 784 + 0,5) / 72,454 ≈ –2,878. Cette valeur étant inférieure au centile 1 de la variable normale standard, soit z[0,01] ≈ –2,326, elle s'écarte donc significativement de 0 et témoigne d'une relation sérieuse, ici une relation non-linéaire, de Y sur X.

Un exemple revisité pour R_{XY}

Le second exemple, celui-ci pour l'indice symétrisé R_{XY}, vient aussi de Laurencelle (2012) et rapporte 15 couples { X, Y }, reproduits aux deux premières colonnes du tableau suivant, et dont le diagramme de corrélation, plus bas, montre une relation décroissante selon X, d'allure vaguement exponentielle. Pour ces données, l'indice « paramétrique » A_S obtient la valeur 0,187, significative au seuil de 1%.

X_i	Y_i	rang X_i	rang Y_i	rang $X_{[i]}$	rang $Y_{[i]}$
1,137	0,101	13	11	12	15
0,824	0,066	8	4	15	14
0,782	0,068	7	5	11	13
0,613	0,099	5	9	8	12
1,014	0,101	10	10	7	9
0,249	0,571	3	13	14	7
1,492	-0,043	15	2	6	5
1,209	0,078	14	6	9	4
0,634	0,08	6	7	5	8
0,596	0,19	4	12	10	10
0,048	0,884	1	15	13	3
1,135	-0,048	12	1	4	1
1,086	0,014	11	3	3	11
0,081	0,786	2	14	2	6
0,889	0,082	9	8	1	2

Le calcul de R_{XY}, un peu plus complexe que R_Y, exige d'abord que, dans l'une et l'autre séries, chaque valeur soit représentée par son rang, de 1 à n, ces rangs apparaissant aux deux colonnes centrales du tableau . Il s'agit ensuite d'obtenir, puis de sommer, les intervalles des concomitantes successives de chaque série de rangs, soit, en langage clair, (1) d'ordonner les deux colonnes de rangs selon Y, obtenant les rangs correspondants de X et apparaissant dans la colonne « rang $X_{[i]}$ », les intervalles |12–15|, |15–11|, etc. étant sommés à 53 et produisant S_X = 53 – (15–1) = 39, puis (2) de refaire l'opération selon X, obtenant les concomitantes de la colonne « rang $Y_{[i]}$ » et les intervalles |15–14|, |14–13|, etc., sommés à 45 et donnant S_Y = 45 – 14 = 31. Finalement, appliquant (4), nous calculons l'indice R_{XY} = (39 + 31) / { 2 · ⌊n(n–2) / 2⌋ } = 70 / 194 ≈ 0,361. Pour n = 15, les valeurs critiques apparaissant au tableau 1 sont 0,464 à 5% et 0,387 à 1%, ce qui aboutit ici aussi à une sanction de significativité. La « corrélation » entre X et Y, quelle que soit sa forme ou son modèle, mérite d'être prise au sérieux.

L'approximation normale, boiteuse pour une série aussi courte que celle-ci à n = 15, peut tout de même être illustrée. Utilisant (6b) à (6d), nous calculons $E(S_X+S_Y)$ ≈ 121,333, $\rho'(n = 15)$ ≈ 5,2 / 17 ≈ 0,306 (l'estimation Monte Carlo fournit ici 0,298), σ^2 ≈ 13·53·16/45·(1 + 0,306) ≈ 310,941 et σ ≈ 17,887. Alors, utilisant encore le modèle (7), nous obtenons z = { (39 + 31) – 121,333 + 0,5 } / 17,887 ≈ –2,842, une valeur significative au seuil de 1%.

Puissance, rangs liés et tutti quanti

Que faut-il penser de ces deux tests non paramétriques dits d'« association non-linéaire » et de leur valeur? En conclusion, nous examinons différentes facettes de ces tests et leur application, en laissant au lecteur le soin d'en juger le mérite.

Puissance. La transformation d'une série de valeurs continues en série de rangs est une catégorisation et, par définition, elle entraîne une perte d'information, partant de puissance. Oui, mais perte de puissance par rapport à quoi? Il n'existe pas de mesure ni de test semblables à référence normale permettant d'apprécier la dépendance non-linéaire entre deux variables, sinon ceux proposés dans Laurencelle (2012) et, peut-être, l'indice η^2 (= somme de carrés-effets / somme de carrés totale) dérivé de l'analyse de variance (cf. p. ex. Winer, 1971), lequel, par contre, exige une catégorisation grossière d'une des deux variables concernées. L'étude de la puissance relative des tests R_Y et R_{XY} par rapport aux tests « normaux » A et A_S reste donc à faire.

L'intérêt d'une telle étude sur la *puissance relative* des tests R_Y et R_{XY} par rapport aux tests A et A_S serait de nous conforter dans l'utilisation des tests non

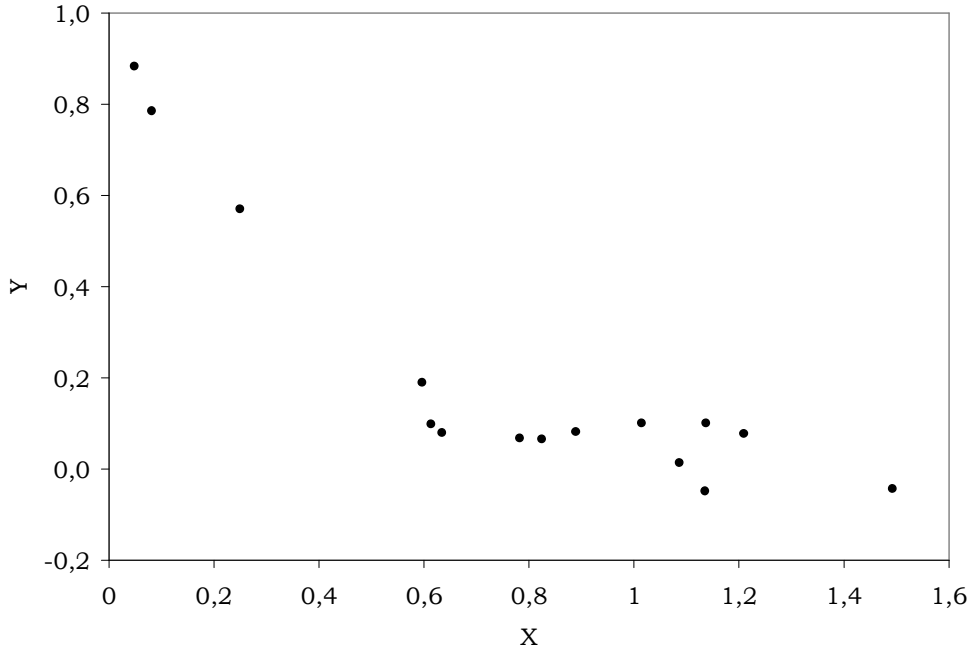

Figure 2 ■ Graphique d'un échantillon de n = 15 données $X_i : Y_i$

paramétriques. Comme l'affirment Kendall et Stuart (1977), « Si nous étudions la puissance [des tests non paramétriques] versus les alternatives considérées dans le contexte de variables normales, nous obtenons un indice de combien nous pouvons perdre en recourant à un test sans référence distributionnelle lorsque les conditions de la théorie normale sont satisfaites (quoique, en réalité, nous n'en sachions vraiment rien). Si cette perte est minime, nous sommes encouragés à sacrifier ce petit surplus d'efficience des méthodes normales au profit du plus large spectre de validité associé aux méthodes sans référence distributionnelle » (p. 500). Ce, d'autant plus, croyons-nous, que, dans le contexte d'une association non-linéaire et d'une relation non modélisée telles que celles envisagées ici, la stipulation du modèle normal est peut-être risquée.

Rangs liés. Même pour des variables continues mais mesurées à nombres de précision finie, l'occurrence de valeurs égales est possible, engendrant alors des rangs égaux, ou « liés ». Le problème des rangs liés est discuté dans la littérature spécialisée (Kendall, 1970; Kendall et Stuart, 1977) et des solutions ad hoc et approximatives se trouvent (p. ex. Lehmann et d'Abrera, 1975; Siegel et Castellan, 1988). Encore une fois, l'étude sur l'impact des rangs liés sur les indices R_Y et R_{XY} mérite d'être faite. Nous faisons néanmoins l'hypothèse que, si le nombre de liens et surtout la grandeur des liens sont proportionnellement petits, leur impact sur la valeur des indices[2] et sur leur performance restera négligeable.

Mérites descriptifs de R_Y et R_{XY} et universalité. Il existe bien sûr d'autres approches pour analyser la dépendance entre deux variables (Drouet-Mari et Kotz, 2001), certaines basées sur la probabilité conditionnelle (Steffensen, 1941; Sylvey, 1964), sur l'élaboration de fonctions bivariées non-linéaires (Scarsini et Venetoulias, 1993) ou sur un réarrangement non-linéaire des données sur matrice graphique (Fisher et Switzer, 1985). Toutes d'un niveau de complexité certain, ces approches accusent aussi la lacune de ne pas proposer de mesure du degré d'interdépendance entre les variables. Nos indices, au contraire, constituent une telle mesure de la dépendance non-linéaire entre X et Y (tout comme, à sa façon, le coefficient η^2 évoqué plus haut), et ils

[2] Dans les cas de certains tests basés sur des rangs, tel le U de Mann-Whitney, l'occurrence de rangs liés (et égaux, obtenus par la moyenne des rangs virtuels attribués arbitrairement aux données égales) affecte d'abord la variance des indices plutôt que leur moyenne. Cependant, l'influence prévisible semble moins importante dans le cas présent, pour lequel les *différences* entre rangs successifs sont calculées.

permettent en outre un test simple de la significativité de cette dépendance. Leur expression sur une échelle de rangs ajoute encore à leur intérêt, leur conférant un domaine d'application quasi universel.

L'histoire décidera de leur sort.

References

Drouet-Mari, D., Kotz S. (2001). *Correlation and dependence*. Singapour : Imperial College Press.

Fisher N. I., Switzer P. (1985). Chi-plots for assessing dependence. *Biometrika*, *72*, 253-265.

Kendall, M. G. (1970). *Rank correlation methods* (4e édition). Londres: Griffin.

Kendall, M.G., Stuart, A. (1977). *The advanced theory of statistics, Vol. 1 :Distribution theory* (4e edition). New York : Macmillan.

Laurencelle, L. (2012). Un indice général d'association entre deux variables continues. Tutorials in Quantitative Methods for Psychology, *8*, 34-43.

Lehmann, E. L., d'Abrera, H. J. M. (1975). *Nonparametrics: statistical methods based on ranks*. Toronto: McGraw-Hill.

Scarsini M, Venetoulias A (1993). Bivariate distributions with nonmonotone dependence structure. *Journal of the American Statistical Association, 88*, 338-344.

Siegel, S., Castellan, N. J. Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2e édition). New York: McGraw-Hill.

Silvey SD (1964). On a measure of association. *Annals of Mathematical Statistics*, *35*, 1157-1166.

Steffensen JF (1941). On the *w* test of dependence between statistical variables. *Skandinavisk Aktuarietidskrift*, *24*, 13-33.

Von Neumann, J., Kent, R. H., Bellinson, H. R., Hart, B. I. (1941a). The mean square successive difference. *Annals of Mathematical Statistics*, *12*, 153-162

Von Neumann, J. (1941b). Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, *12*, 367-395.

Winer, B. J. (1971). *Statistical principles in experimental design* (2e édition). New York: McGraw-Hill.

Citation

Laurencelle, L. (2015). A general non-linear index of association between two continuous rank-order variables. *The Quantitative Methods for Psychology*, *11* (1), 1-7.

Received: 11/03/14 ~ Accepted: 11/12/14