

Oui, mais est-ce que votre analyse en réseau est *implicative* ?

Corentin Gonthier

univ-nantes.fr/corentin-gonthier



institut
universitaire
de France

Bruno Dauvier

centrepsyche-amu.fr/bruno-dauvier/





<https://osf.io/c5h78/>

```
#####
# Examples
#####

# Generate a dataset containing two variables A and B with an implicative shape
testDataset <- implicationSimulatedData(n=200, corAB=0.50, implicationThreshold=0.90, winsorize=0.001)

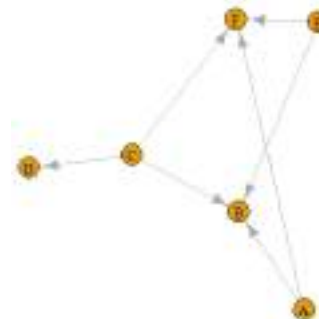
# Compute npsi index of implication for both a=>b and b=>a, as well as the corresponding p-values with 1000 samples
# Notes: 10.000 samples is preferable for actual use ; normalize is set to FALSE because the dataset generated above is already normalized and winsorize
npsi <- ASI.npsi(testDataset$A.norm, testDataset$B.norm, normalize=F, winsorize=0.001, p=TRUE, samples=1000)

# Plot A and B and report the implication values and p-values
plot(testDataset$A.norm, testDataset$B.norm, pch=20, xlab="A", ylab="B")
legend(x="bottomright", legend=c(paste("npsi.a=>b = ", round(npsi$npsi.ab, 3), ", p=", round(npsi$npsi.ab.p, 3), sep=""), paste("npsi.b=>a = ", round(npsi$npsi.ba, 3), ", p=", round(npsi$npsi.ba.p, 3), sep="")), bty="n")

# Generate a longer dataset containing six variables with implicative relationships
testDataset1 <- implicationSimulatedData(n=200, corAB=0.50, implicationThreshold=0.90, winsorize=0.001)[,2:3]
testDataset2 <- implicationSimulatedData(n=200, corAB=0.40, implicationThreshold=0.80, winsorize=0.001)[,2:3]
testDataset3 <- implicationSimulatedData(n=200, corAB=0.30, implicationThreshold=0.70, winsorize=0.001)[,2:3]
fullDataset <- data.frame(cbind(testDataset1, testDataset2, testDataset3))
colnames(fullDataset) <- c("A","B","C","D","E","F")

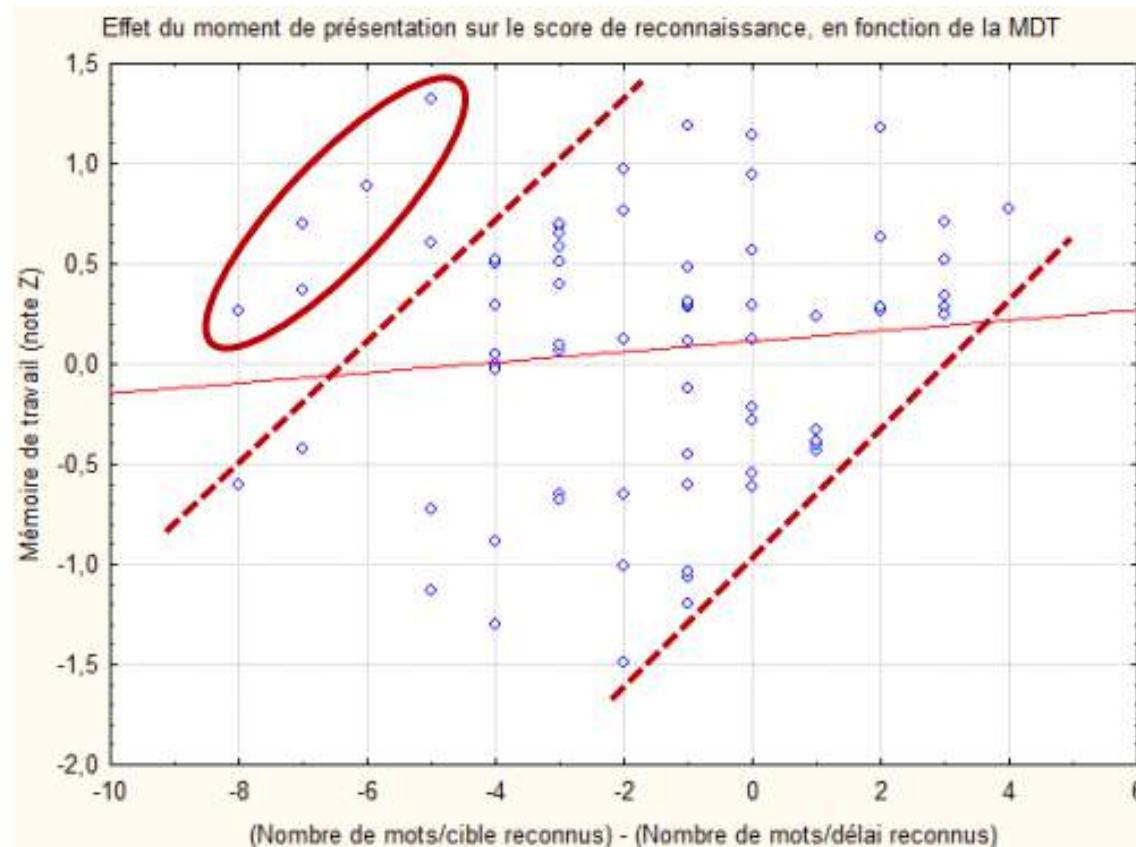
# Draw a directed implicative network, showing only relations that are significant at p<.05
# Notes: 10.000 samples is preferable for actual use ; normalize is set to FALSE because the dataset generated above is already normalized and winsorize
# Can take a minute to run
generateImplicationGraph(dataset=fullDataset, firstColumn=1, normalize=F, winsorize=0.001, samples=1000, threshold=0.05)
```

	A	B	C	D	E	F
A	0	0.659	0	0.000	0	0.224
B	0	0.000	0	0.000	0	0.000
C	0	0.567	0	0.582	0	0.308
D	0	0.000	0	0.000	0	0.000
E	0	0.459	0	0.000	0	0.403
F	0	0.000	0	0.000	0	0.000



L'analyse implicative : pourquoi ???

Mais pourquoi ?



Conclusion

Les résultats vont dans le sens d'un traitement plus réactif pour les participants à faible MDT. En revanche, tout se passe comme si les participants à forte MDT pouvaient utiliser les deux types de traitements. Il semble donc qu'il n'y ait pas d'équivalence stricte entre MDT et mécanisme de contrôle utilisé : on observe plutôt un lien d'implication.

(Gonthier et al., 2012)

Mais pourquoi ?

12/03/2025 17:48

Différences et variabilités en psychologie - L'augmentation des capacités en mémoire de travail est-elle une condition nécessa...

L'augmentation des capacités en mémoire de travail est-elle une condition nécessaire, bien que non suffisante, au développement de l'intelligence ?

Christine Bailleux, Bruno Dauvier et Patrick Perret

p. 37-62

TEXTE INTÉGRAL

Introduction

- 1 Les chercheurs en psychologie cognitive ont longtemps étudié des processus cognitifs de hauts niveaux en négligeant la question de leurs liens avec l'intelligence. Ils envisagent maintenant les capacités de la mémoire de travail, et notamment l'intégration des relations, comme l'un des principaux facteurs qui sous-tendent les différences individuelles d'intelligence générale (Conway, Kane, & Engle, 2003 ; Hornung, Brunner, Reuter & Martin, 2011). Cependant, la mesure qu'ils proposent de la capacité en mémoire de travail n'est pas univoque et peu d'auteurs, mis à part Cowan (1999), intègrent ces



(Bailleux et al. 2019)

Mais pourquoi ?

Mais aussi : une question classique de la psychométrie

1. Accepts country in Country	2. Accepts country in Town	3. Accepts country in Neighborhood	4. Accepts country Next Door	5. Accepts an country as Spouse	Possible Scale Score
0	0	0	0	0	0
1	0	0	0	0	1
1	1	0	0	0	2
1	1	1	0	0	3
1	1	1	1	0	4
1	1	1	1	1	5

Echelle de Guttman :
homogénéité par implication

Yvanick Noel
Analyse implicative et modèle de Rasch

Motivation
Indices d'association sur données binaires
L'analyse statistique implicative
Un modèle dimensionnel de l'implication
Considérations

Constructive
Lien au modèle de Rasch
Avec des données ordonnées (idéaux, TACD)

Notion de distance implicative

- Dans le cas Rasch, l'implication statistique augmente **avec la distance latente**.
- Sur 3 items ordonnés X_1 , X_2 et X_3 , dans le cas parfait, on a (**additivité des distances**) : $(\delta_3 - \delta_1) = (\delta_3 - \delta_2) + (\delta_2 - \delta_1)$.
- Sur les indices $|\ln i|$ observés, en valeurs absolues, on devrait pouvoir retrouver une **additivité approximative** sur les distances inter-items d_{ij} qui s'en déduisent :

$$d_{13}^{(1)} \approx d_{12}^{(1)} + d_{23}^{(1)}$$

Modèle de Rasch :
quasi-implication (implicite)

Le groupe de Régis Gras

Le groupe de Régis Gras

- Régis Gras (1933-2022)
- Professeur émérite à l'école polytechnique de Nantes
- Didactique des mathématiques, métaphysique (?)



FRACTALITE DU RYTHME DU TEMPS SUR L'ARBRE DE VIE

Régis Gras¹, Raphael Conturier², Saddy Ag Almouloud³

Résumé

Le texte que nous présentons ci-dessous se charpente selon plusieurs facettes renvoyant toutes à la notion de fractalité, concept mathématique géométrique imaginé par Mandelbrot. Nous avons fréquenté ce concept en A.S.I. lors de l'étude des représentations graphiques des relations causales (graphe et arbre). Cet article vise à renforcer la familiarisation du lecteur avec cette géométrie en lui montrant un autre champ d'application et sa puissance créative. Nous partons d'une question vive qui perturbe chacun d'entre nous et de façon accentuée au fil de la vie, à savoir la différence numérique entre le temps réel, c'est-à-dire celui de l'horloge et le temps perçu par un sujet existant. Moyennant deux hypothèses plausibles sur la relation entre ces deux mesures, l'une physique, l'autre psychologique, nous modélisons une formule générique reliant ces deux notions. Celles-ci constituent une distorsion, autrement dit un biais cognitif. Nous montrons qu'il s'agit d'un phénomène fractal plongé dans l'espace-temps de la relativité restreinte. En effet, il apparaît que la représentation graphique logarithmique des covariations de ces deux notions est linéaire. Nous savons que cette dernière propriété, comme nous l'avons observé par ailleurs avec les images graphiques de l'A.S.I., est caractéristique d'un phénomène fractal.

Mots-clés : biais cognitif, temps réel, temps ressenti, temps perçu, relativité, espace-temps, échelle logarithmique, fractal.

(Gras et al. 2021)

Le groupe de Régis Gras

- Pascale Kuntz
 - Professeure des Universités à l'école polytechnique de Nantes
 - M'a aidé à aborder le sujet

- Jean-Claude Régnier
 - Professeur des universités à Lyon 2
 - M'a aidé à comprendre les sorties logicielles

- Raphaël Couturier
 - Professeur des universités à l'Université Franche-Comté
 - A rendu accessible le code de RCHIC



La logique implicative

La logique implicative

Corrélation entre A et B

- $A \leftrightarrow B$
- Plus A augmente, plus B augmente
- Plus B augmente, plus A augmente
- A se produit généralement avec B
- B se produit généralement avec A

B	4	22
non-B	17	3
	non-A	A

😊	pas trop	souvent
😞	souvent	pas trop
	💧	☀️

A implique B

- $A \rightarrow B$
- Quand A est élevé, B l'est aussi
- Quand B est élevé, A peut être faible
- A se produit toujours avec B
- B se produit parfois avec A

B	19	22
non-B	17	3
	non-A	A

😊	souvent	souvent
😞	souvent	pas trop
	💣	📄

La logique implicative

L'implication $\rightarrow$ est une moitié de corrélation $\leftrightarrow$: une relation asymétrique

B	19	22
non-B	17	3
	non-A	A

Remarque : la littérature parle de
"quasi-implication"

ce n'est pas une implication causale **parfaite**
("quand papier, TOUJOURS CONTENT")

😊	souvent	souvent
😞	souvent	pas trop
		

La logique implicative

➤ Le piège de la causalité

measures proposed in the literature, it does not take into account the contrapositive $\bar{b} \Rightarrow \bar{a}$ which could allow to reinforce the affirmation of the good quality of the implicative relationship between a and b , and the capacity to estimate the causality between the variables.

2.2.3 Causalité

Nous dirons qu'il y a relation de type causal⁷ de a sur b si nous observons les conditions suivantes : d'une part, b se produit quand ou dès que a se produit, et d'autre part, b ne se produit pas lorsque ou alors que a ne se produit pas. Autrement dit (TAB. 2) $(a(x) \wedge \bar{b}(x))$ est faux dans toute circonstance x .

1	2	3	4
$a(x)$	$b(x)$	$\text{non-}b(x)$	$a(x) \wedge \text{non-}b(x)$
v	v	f	f
f	f	v	f

TAB. 2 : Tables de vérité

establish because of the exceptions. Consequently, it is necessary the dataset to extract rules reliable enough to conjecture causal which structure the population. At the descriptive level, they et a certain stability in the structuration. And, at the predictive low to make assumptions. However, the rule mining processes

😊	souvent	souvent
😞	souvent	pas trop
	💣	📄



Causal ou pas causal ?

IMPLICATION ENTROPIQUE ET CAUSALITE

Régis GRAS¹, Raphaël COUTURIER²

TITLE

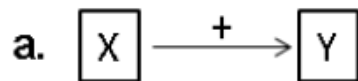
Entropy implication and causality

RESUME

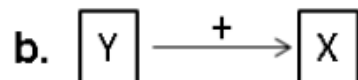
La méthode d'analyse de données, l'Analyse Statistique Implicative (A.S.I.), a pour objectif, à partir du croisement d'une population et de variables, l'extraction et l'étude de certaines relations d'association, appelées règles, de type implicatif (donc non-symétriques). Une règle statistique, l'entropie d'implication, permet d'attribuer un indicateur numérique

La logique implicative

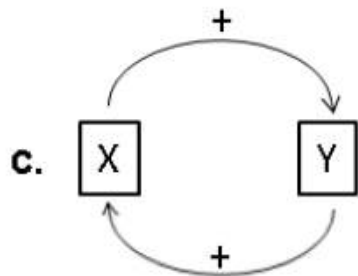
- Exemple : chez l'enfant, on trouve une faible corrélation entre nombre d'heures moyen devant la télévision (X) et agressivité (Y).



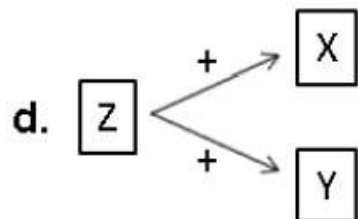
Explication a. La télévision rend agressif.



Explication b. Les enfants agressifs ont plus tendance à regarder des dessins animés violents à la télévision.



Explication c. La télévision rend agressif, et plus on est agressif, plus on a plus envie de regarder des dessins animés violents à la télévision.



Explication d. Les enfants issus d'un milieu défavorisé passent plus de temps devant la télévision et sont également plus agressifs.

La logique implicative

- A est une condition suffisante mais non nécessaire de B (si papier alors content, mais content possible même si pas papier)
- B est une condition nécessaire mais non suffisante de A (content requis pour que papier, mais si content pas forcément papier)
- A et B forment une boucle causale asymétrique (B permet A qui augmente la probabilité de B qui...)
- C détermine à la fois A et B, avec une contrainte plus importante sur A

😊	souvent	souvent
😞	souvent	pas trop
	💣	📄

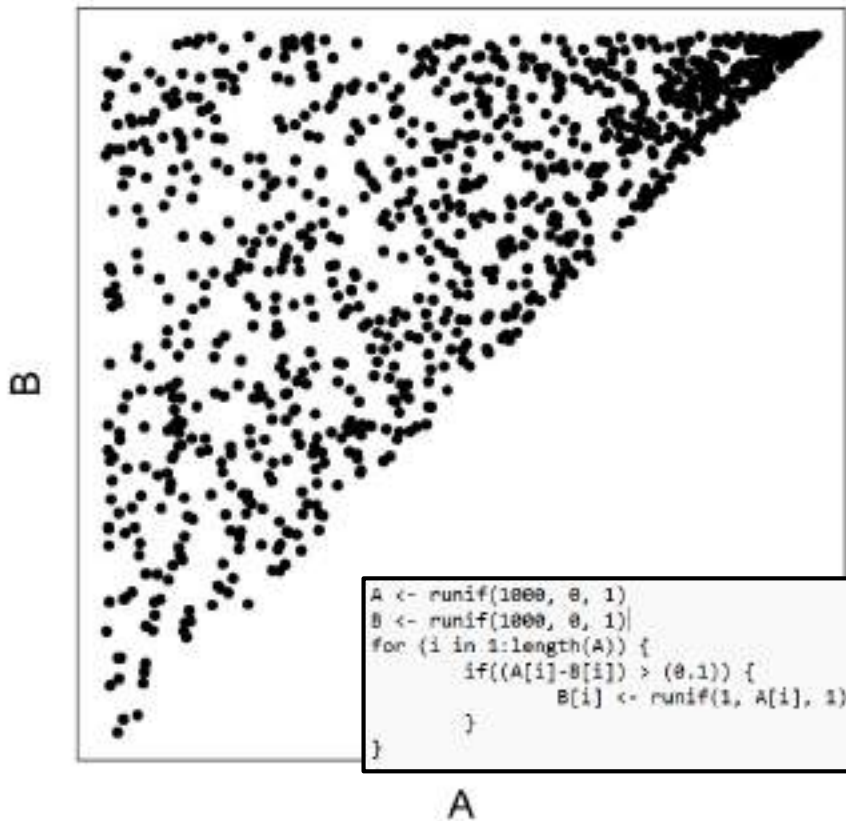


Causal ou pas causal ?

La logique implicative

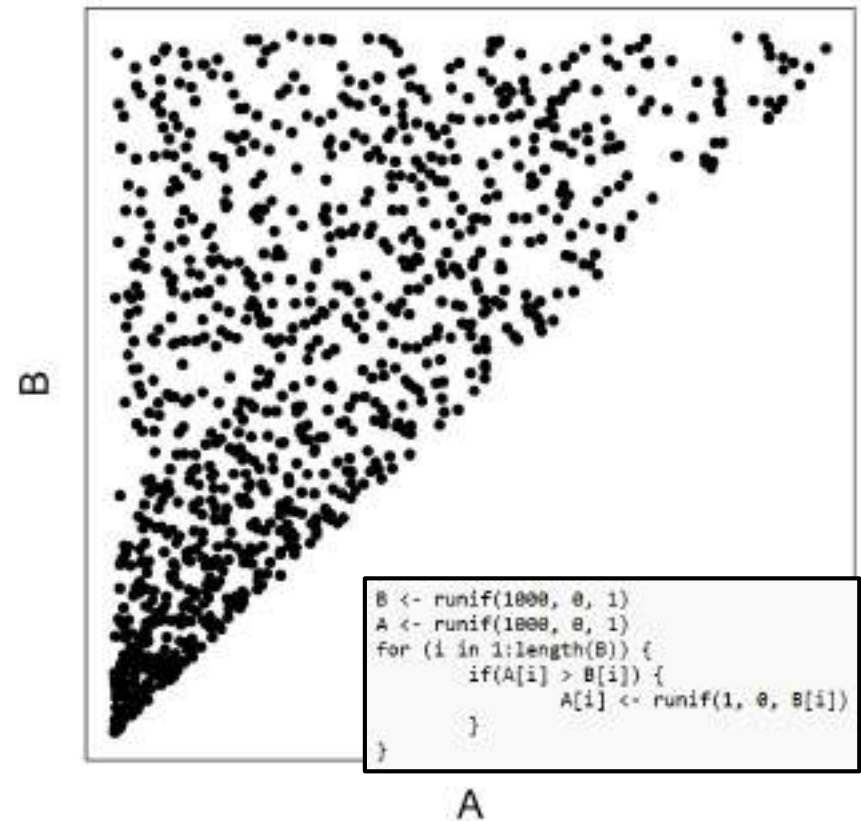
A est une cause suffisante non-nécessaire de B

Redraw any B less than A



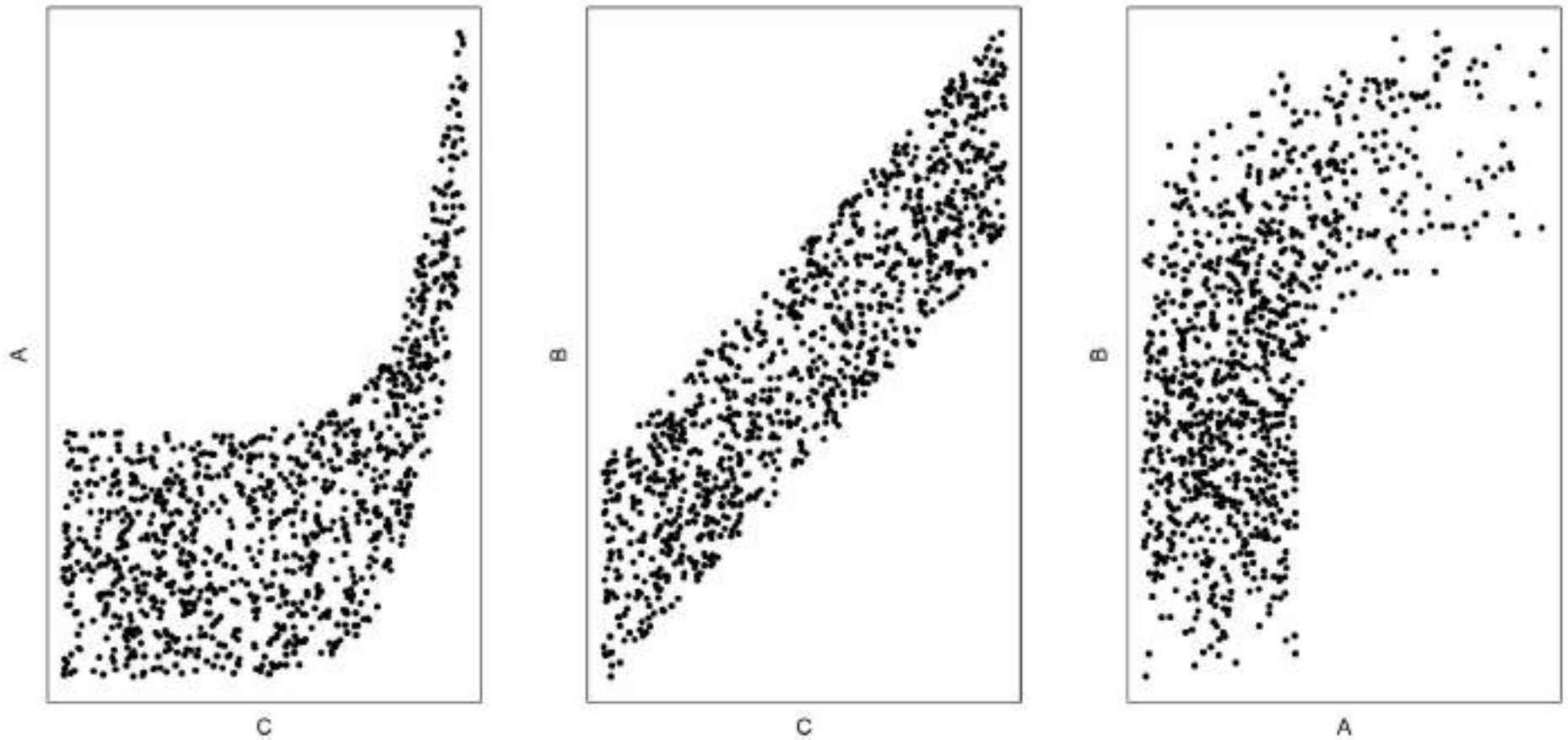
B est une cause nécessaire non-suffisante de A

Redraw any A more than B



La logique implicative

A implique B parce que C contraint plus fortement A que B



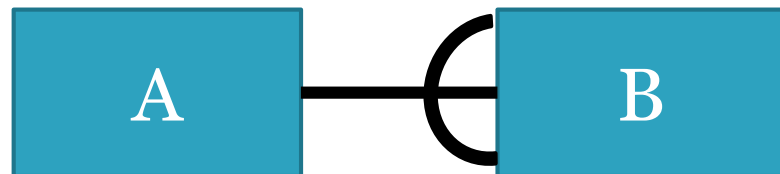
```
C <- runif(1000, 0.25, 1.25)
A <- C^10 + runif(1000, 0, 5)
B <- 2*C + runif(1000, 0, 1)
```

La logique implicative

- A est une condition **suffisante** **mais** **non nécessaire** de B
(si papier alors content, mais content possible même si pas papier)
- B est une condition **nécessaire** **mais** **non suffisante** de A
(content requis pour que papier, mais si content pas forcément papier)
- A et B forment une boucle causale **asymétrique**
(B permet A qui augmente la probabilité de B qui...)
- C détermine à la fois A et B, avec une contrainte **plus importante** sur A

→ L'analyse implicative teste des relations (causales) **asymétriques**

Plutôt $A \in B$ que $A \rightarrow B$...



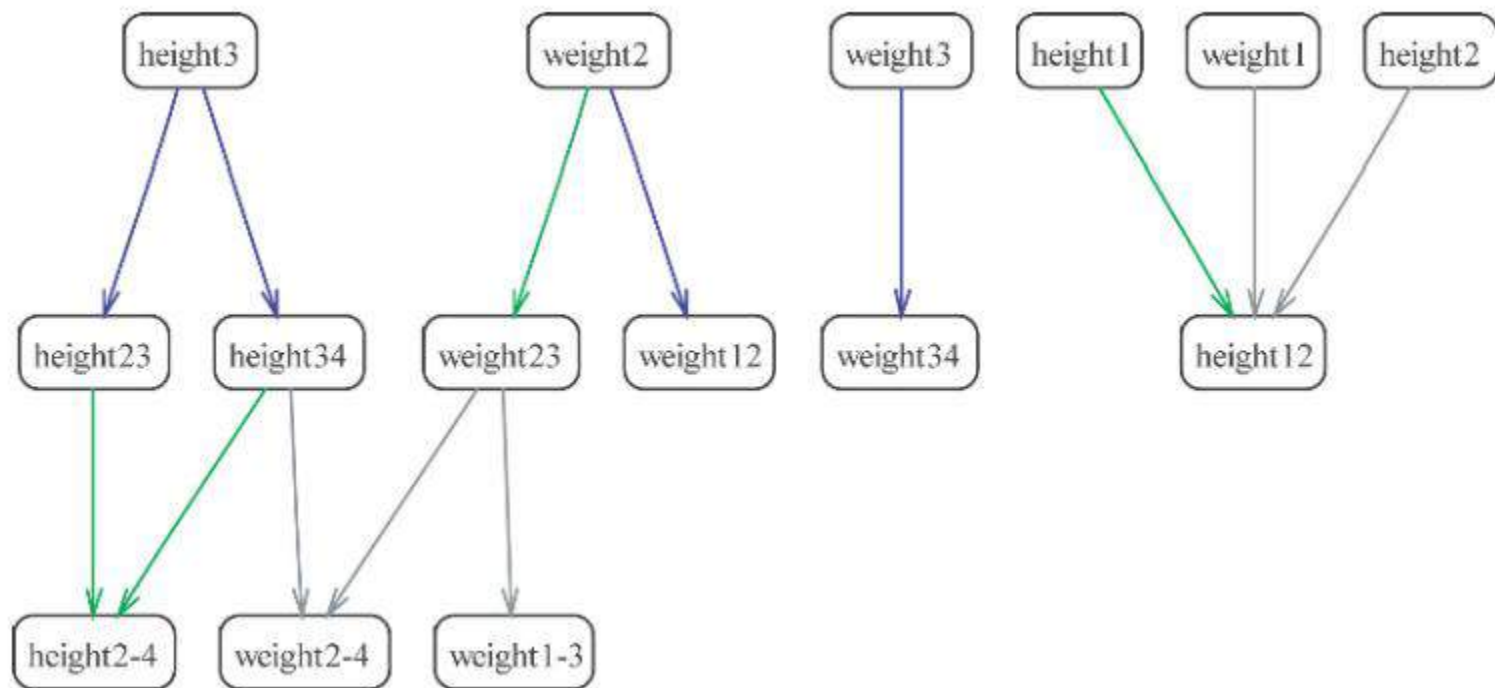
Extension du cas binaire au cas numérique

Du binaire au numérique

Une première option à écarter :

binariser en recodant en intervalles

- arbitrairement, automatiquement (par analyse en clusters)

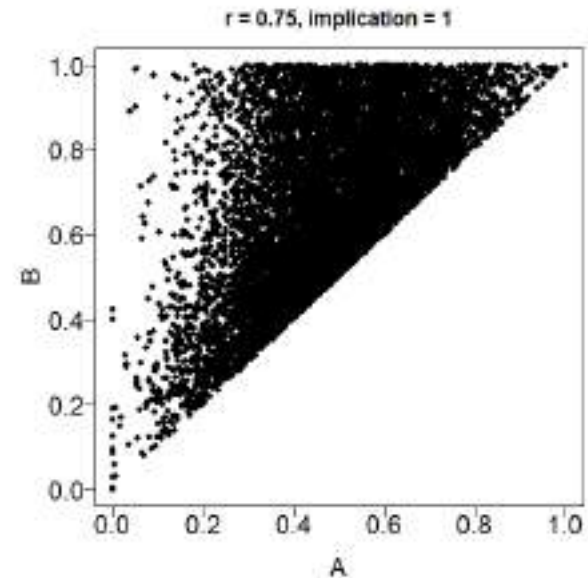
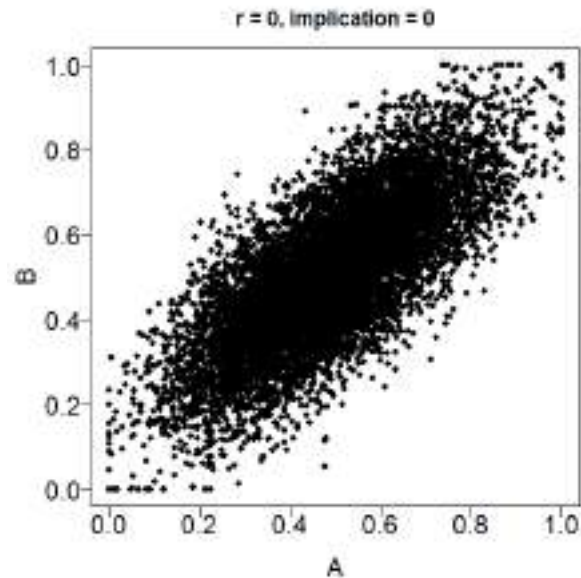


Du binaire au numérique

En réalité l'extension de principe est assez simple :

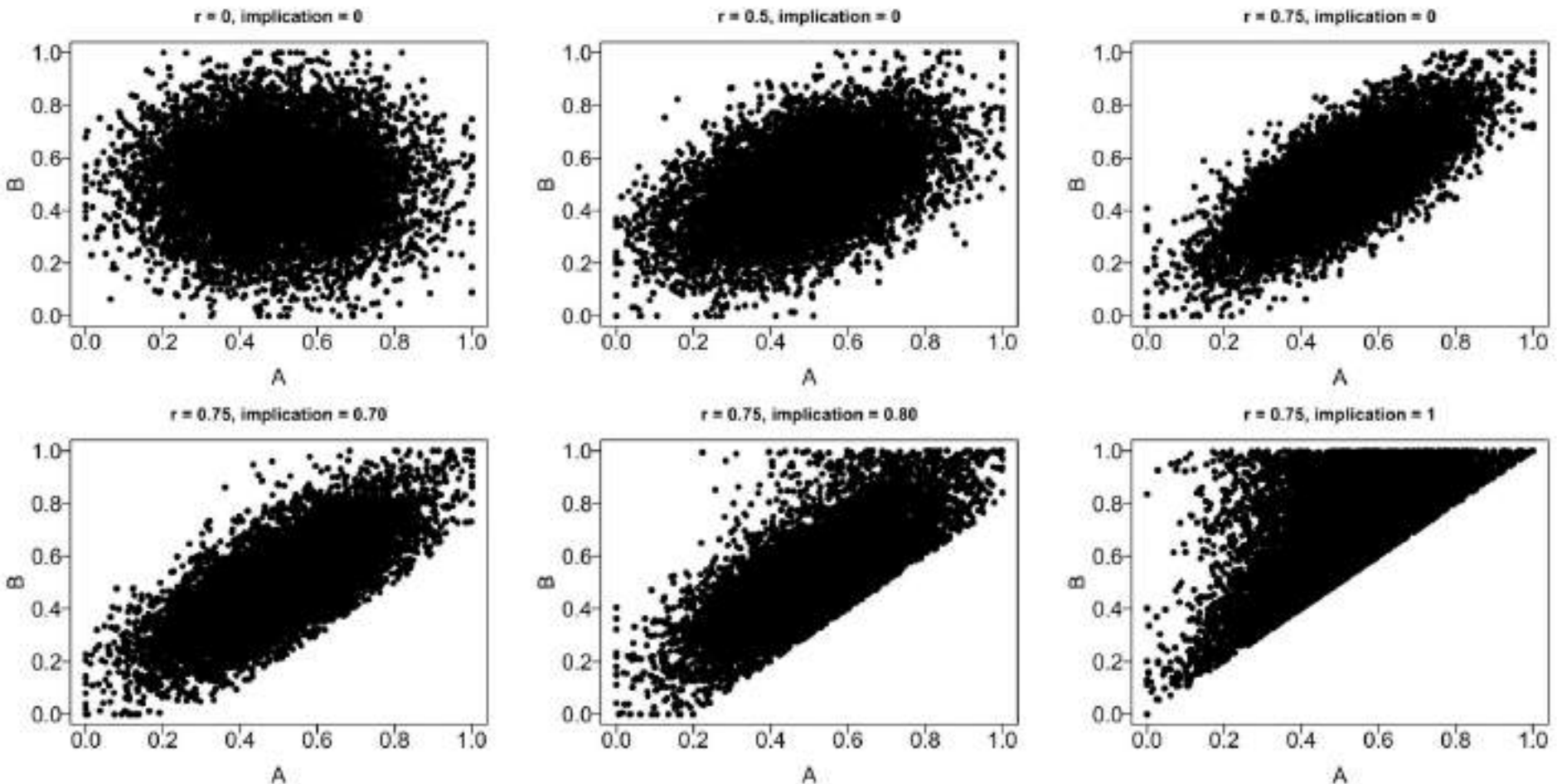
B	4	22
non-B	17	3
	non-A	A

B	19	22
non-B	17	3
	non-A	A



Du binaire au numérique

Générer des données "plausibles" :
recodage des valeurs de B si $(A-B) > (1-\text{implication})$



Du binaire au numérique

- Une variable binaire A/non-A peut être recodée en 0/1 (nombre de A == somme de A=1)
- Une variable non-binaire A peut être recodée comme la somme des valeurs de A

En fait, une suggestion de I.C. LERMAN confortant notre propre approche, permet d'interpréter ces variables en terme de *variables fréquentielles* définies de la façon suivante :

Soit $(a, b) \in V^2$ et $i \in E$, α (resp. β) la valeur maximum de a (resp. b) sur E , α_i (resp. β_i) la *valeur observée* de a (resp. b) sur i . Posons :

$$\alpha'_i = \frac{\alpha_i}{\alpha}, \beta'_i = \frac{\beta_i}{\beta}, \nu_a = \sum_{i \in E} \alpha'_i, \nu_b = \sum_{i \in E} \beta'_i, \nu_{a \wedge b} = \sum_{i \in E} \alpha'_i (1 - \beta'_i) \text{ et } \nu_{\bar{a}} = n - \nu_a.$$

Ces valeurs pour $\alpha = \beta = 1$, α_i et $\beta_i = 1$ ou 0 étendent alors le cas binaire où les effectifs fréquentiels sont n_a , n_b et $n_{a \wedge b}$. En effet, si 1_A et 1_B sont les fonctions indicatrices des sous-ensembles d'individus possédant respectivement les caractères a et b , l'effectif $n_{a \wedge b}$ s'écrit :

$$n_{a \wedge b} = \sum_{i \in E} 1_A(i) [1 - 1_B(i)]$$

(Gras & Larher 1992)

- Cette approche donne une approximation satisfaisante des indices de l'ASI...
- ...mais de valeur inférieure dans le cas usuel d'un facteur de variance > 2

Soit a (resp. b) une variable de moyenne empirique m_a (resp. m_b) et de variance empirique ν_a (resp. ν_b).

$$\tilde{q}(a, b) = \frac{\sum_{i \in E} \alpha_i (1 - \beta_i) - m_a (1 - m_b)}{\sqrt{\frac{(\nu_a + m_a^2)(\nu_b + (1 - m_b)^2)}{n}}}$$

est le coefficient de propension de a vers b .

a vers b ,

• $\Phi(-\tilde{q}(a, b))$ est l'intensité de propension de a vers b .

Une variable binaire est égale à son carré, et donc si a et b sont binaires, avec les notations ci-dessus : $\nu_a + m_a^2 = m_a$ et $\nu_b + (1 - m_b)^2 = m_b$ d'où

$$\tilde{q}(a, b) = \frac{s - m_a(1 - m_b)}{\sqrt{m_a(1 - m_b)}}$$

Dans ce cas, $\tilde{q}(a, b)$ coïncide avec le coefficient $q(a, b)$ utilisé en analyse implicite.

Dans le cas binaire, l'intensité de propension est donc égale à l'intensité d'implication $\Phi(-q(a, b))$ introduite par Gras (1979) et Lerman, Gras, Rostram, (1981).

Conséquence : le Tableau 1 donne l'encadrement de λ pour les valeurs de k de 2 à 4, puis des encadrements de l'intensité d'implication pour une intensité de propension valant 0,9 puis 0,95. On constate que pour ces valeurs typiques d'intensités significatives, des valeurs du facteur de variance supérieures ou égales à 2 donnent une intensité d'implication notablement inférieure à l'intensité de propension. Il faudra

(Lagrange 1995)

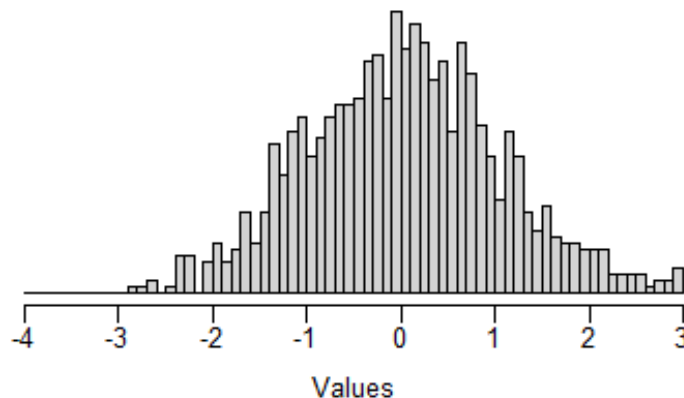
Du binaire au numérique

- Une contrainte : que les données soient bornées entre 0 et 1

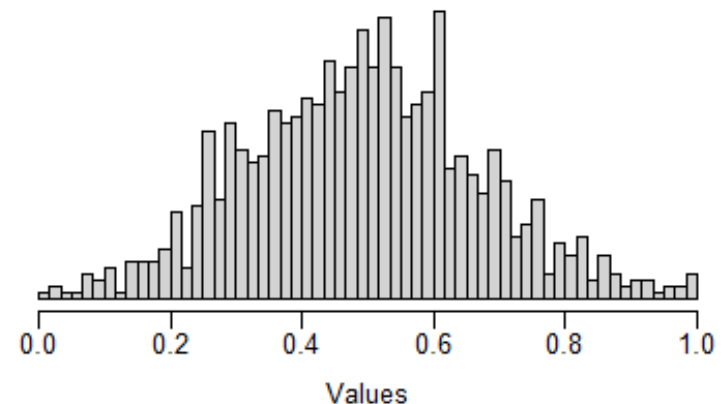
➔ renormalisation obligatoire

$$X = [(X - \min(X)) / (\max(X) - \min(X))]$$

Mean of A = 0.03 | Variance of A = 1.07



Mean of A = 0.49 | Variance of A = 0.03



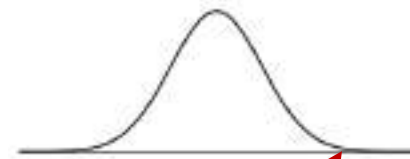
Statistiques pour une logique implicative

Statistiques implicatives

- Stratégie 1 : le coefficient q et l'intensité d'implication ϕ
- Indice "historique" de l'ASI

B	20	20
non-B	20	5
	non-A	A

l'indice de Loevinger, (Loevinger 1947), l'aîné semble-t-il des indices d'implication de a sur b , $H(a, b) = 1 - \frac{n \cdot n_{a \wedge \bar{b}}}{n_a n_{\bar{b}}}$



$$q(a, \bar{b}) = \frac{n_{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

$$\varphi(a, b) = 1 - \Pr(Q(a, \bar{b}) \leq q(a, \bar{b})) = \frac{1}{\sqrt{2\pi}} \int_{q(a, \bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

- "1 - la probabilité (x 100) d'obtenir aussi peu de contre-exemples à la règle d'implication si les variables étaient indépendantes"
- "La surprise face à la rareté des contre-exemples à la règle implicative"

Statistiques implicatives

- Stratégie 1 : le coefficient q et l'intensité d'implication ϕ
- Indice "historique" de l'ASI

$$q(a, \bar{b}) = \frac{n_{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

- Avec des données numériques :

$$\tilde{q}(a, \bar{b}) = \frac{\sum_{i \in P} a_i(1 - b_i) - m_a(1 - m_b)}{\sqrt{\frac{(\nu_a + m_a^2)(\nu_b + (1 - m_b)^2)}{n}}}$$

= - 1 x cov(A, B)

Une fonction de l'écart entre les moyennes et des variances de A et B

Statistiques implicatives

- Stratégie 2 : les indices de confiance
- Des classiques de la littérature

Les valeurs de la confiance relative respectivement
aux règles $a \Rightarrow b$ et $\bar{b} \Rightarrow \bar{a}$ sont

$$C_1(a,b) = \text{Fr}[Y | X] = ([\text{card } X \cap Y]) / (\text{card } X) = \frac{n_{a \wedge b} / n}{n_a / n} = \frac{n_{a \wedge b}}{n_a}$$

$$C_2(\bar{b}, \bar{a}) = \text{Fr}[\bar{X} | \bar{Y}] = ([\text{card } \bar{X} \cap \bar{Y}]) / (\text{card } \bar{Y}) = \frac{n_{\bar{a} \wedge \bar{b}} / n}{n_{\bar{b}} / n} = \frac{n_{\bar{a} \wedge \bar{b}}}{n_{\bar{b}}}$$

B		
		A & B
non-B		A & non-B
	non-A	A

B		
non-B	non-A & non-B	
		A & non-B
	non-A	A

Statistiques implicatives

- Stratégie 2 : les indices de confiance
- Des classiques de la littérature

Les valeurs de la confiance relative respectivement
aux règles $a \Rightarrow b$ et $\bar{b} \Rightarrow \bar{a}$ sont

$$C_1(a,b) = \text{Fr} [Y | X] = ([\text{card } X \cap Y]) / (\text{card } X) = \frac{n_{a \wedge b}/n}{n_a/n} = \frac{n_{a \wedge b}}{n_a}$$

$$C_2(\bar{b}, \bar{a}) = \text{Fr} [\bar{X} | \bar{Y}] = ([\text{card } \bar{X} \cap \bar{Y}]) / (\text{card } \bar{Y}) = \frac{n_{\bar{a} \wedge \bar{b}}/n}{n_{\bar{b}}/n} = \frac{n_{\bar{a} \wedge \bar{b}}}{n_{\bar{b}}}$$

- Avec des données numériques :
 - $C1 = \Sigma(A \times B) / \Sigma(A)$
 - $C2 = \Sigma(1-A \times 1-B) / \Sigma(1-B)$

Statistiques implicatives

Déclinés en version entropique par Gras et al.

Posons $\alpha = n_a / n$ (resp. $\beta = n_{\neg b} / n$) la fréquence de a (resp. non b) sur la population. Selon les contraintes précédentes, les deux "ajustements" de l'entropie significatifs des qualités respectives de l'implication et de sa contraposée peuvent donc s'écrire en fonction de la fréquence $t = n_{a \rightarrow b} / n$ de contre-exemples de la façon suivante $\square$

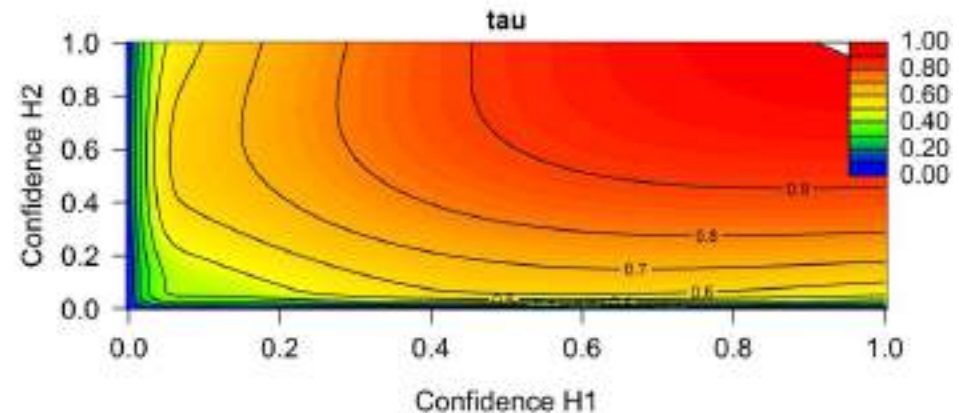
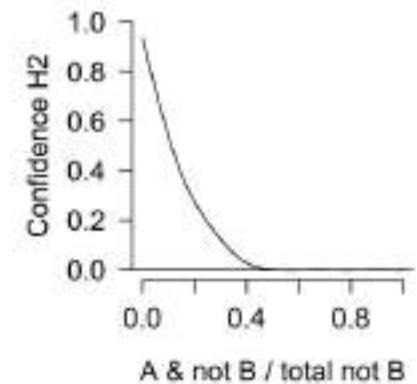
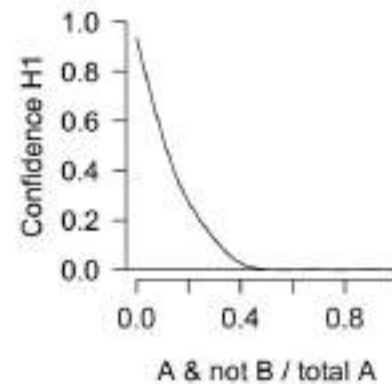
$$h_1(t) = H(b/a) = - \left(1 - \frac{t}{\alpha}\right) \log_2 \left(1 - \frac{t}{\alpha}\right) - \frac{t}{\alpha} \log_2 \left(\frac{t}{\alpha}\right) \quad \text{si } t \in \left[0, \frac{\alpha}{2}\right]$$

$$h_1(t) = 0 \quad \text{si } t \in \left[\frac{\alpha}{2}, \alpha\right]$$

$$h_2(t) = H(\neg a / \neg b) = - \left(1 - \frac{t}{\beta}\right) \log_2 \left(1 - \frac{t}{\beta}\right) - \frac{t}{\beta} \log_2 \left(\frac{t}{\beta}\right) \quad \text{si } t \in \left[0, \frac{\beta}{2}\right]$$

$$h_2(t) = 0 \quad \text{si } t \in \left[\frac{\beta}{2}, \beta\right]$$

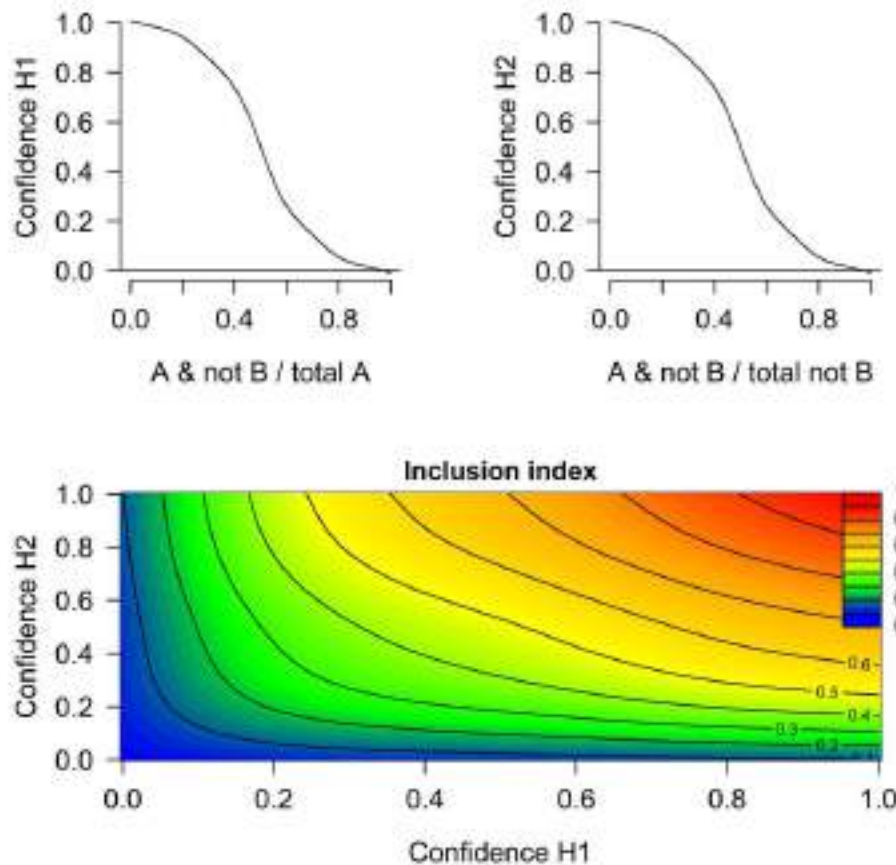
$$\tau(a, b) = \left(1 - h_1^2(t)\right) \cdot \left(1 - h_2^2(t)\right)^{1/4}$$



(Gras et al. 2001)

Statistiques implicatives

Déclinés en version entropique par Gras et al.



$$H(b/a) = -\frac{n_{a \wedge b}}{n_a} \log_2 \frac{n_{a \wedge b}}{n_a} - \frac{n_{a \wedge \bar{b}}}{n_a} \log_2 \frac{n_{a \wedge \bar{b}}}{n_a}$$

$$H(\bar{a}/\bar{b}) = -\frac{n_{\bar{a} \wedge \bar{b}}}{n_{\bar{b}}} \log_2 \frac{n_{\bar{a} \wedge \bar{b}}}{n_{\bar{b}}} - \frac{n_{\bar{a} \wedge b}}{n_{\bar{b}}} \log_2 \frac{n_{\bar{a} \wedge b}}{n_{\bar{b}}}$$

$$i(a, b) = \left(\left[1 - h_1(t) \right] \left[1 - h_2(t) \right] \right)^{\frac{1}{2}}$$

$$\Psi(a, b) = \left(1 - \frac{1}{2\sqrt{n}} \right) i(a, b)$$

(Gras & Couturier 2012)

Statistiques implicatives

Yvonnick Noël	Analyse implicative et modèle de Rasch
Motivation Indices d'association sur données binaires L'analyse statistique implicative Un modèle dimensionnel de l'implication Conclusions	Construction Lien au modèle de Rasch Jeux de données réelles (Sciences, TACIT)
Statistique d'implication	

- Au final, les deux attendus sont réunis dans la **statistique produit** (indice « iota ») :

$$\iota_{A \rightarrow B} = o_{b|a} \times o_{\bar{a}|\bar{b}} = \frac{\pi_{ab}}{\pi_{a\bar{b}}} \times \frac{\pi_{\bar{a}\bar{b}}}{\pi_{a\bar{b}}} = \frac{\pi_{ab}\pi_{\bar{a}\bar{b}}}{\pi_{a\bar{b}}^2}$$

- Propriétés :
 - il varie entre 0 et $+\infty$ (implication parfaite, $\pi_{a\bar{b}} = 0$).
 - Dans ce dernier cas, **l'indice ι est infini** (discussion plus loin).
 - Il vaut 1 en situation d'**incertitude totale** ($\pi_{ab} = \pi_{a\bar{b}}$ et $\pi_{\bar{a}\bar{b}} = \pi_{a\bar{b}}$).

(Noël 2017)

Statistiques implicatives

➤ Stratégie 3 : Frankenstein



Posons $\alpha = n_a / n$ (resp. $\beta = n_{\neg b} / n$) la fréquence de a (resp. non b) sur la population. Selon les contraintes précédentes, les deux "ajustements" de l'entropie significatifs des qualités respectives de l'implication et de sa contraposée peuvent donc s'écrire en fonction de la fréquence $t = n_{a \Rightarrow b} / n$ de contre-exemples de la façon suivante $\square$

$$h_1(t) = H(b/a) = - \left(1 - \frac{t}{\alpha}\right) \log_2 \left(1 - \frac{t}{\alpha}\right) - \frac{t}{\alpha} \log_2 \left(\frac{t}{\alpha}\right) \quad \text{si } t \in \left[0, \frac{\alpha}{2}\right]$$

$$h_1(t) = 0 \quad \text{si } t \in \left[\frac{\alpha}{2}, \alpha\right]$$

$$h_2(t) = H(\neg a / \neg b) = - \left(1 - \frac{t}{\beta}\right) \log_2 \left(1 - \frac{t}{\beta}\right) - \frac{t}{\beta} \log_2 \left(\frac{t}{\beta}\right) \quad \text{si } t \in \left[0, \frac{\beta}{2}\right]$$

$$h_2(t) = 0 \quad \text{si } t \in \left[\frac{\beta}{2}, \beta\right]$$

DÉFINITION 11. L'indice d'inclusion de I_a , support de a , dans I_b , support de b , est le nombre qui intègre l'information délivrée par la réalisation d'un faible nombre de contre-exemples, d'une part à la règle $a \Rightarrow b$ et, d'autre part, à la règle $\neg b \Rightarrow \neg a$.

$$\tau(a, b) = \left(1 - h_1^2(t)\right) \cdot \left(1 - h_2^2(t)\right)^{1/4}$$

L'intensité entropique de la règle $a \Rightarrow b$ est définie par $\psi(a, b) = (\varphi(a, b) \cdot \tau(a, b))^{1/2}$ où $\varphi(a, b)$ est l'intensité d'implication et $\tau(a, b)$ l'indice d'inclusion.

(Gras et al. 2001)

Statistiques implicatives

➤ Stratégie 3 : Frankenstein



6.2 La corbeille des mariés.

Définition 4

On appelle **implifiance**⁷ la mesure de l'implication statistique qui prend en compte l'implication directe et sa contraposée ainsi que la confiance en chacune de ces deux formes inclusives. Sa valeur est :

$$\Phi(a,b) = \varphi(a,b) \cdot [C_1(a,b) \cdot C_2(\bar{b}, \bar{a})]^{1/4}$$

Ainsi, même si les covariations de l'intensité d'implication et la confiance ne sont pas anarchiques, si elles n'obéissent pas ensemble à une référence en termes de probabilité, même si nous nous trouvons devant le choix arbitraire d'une définition composite d'où est partiellement exclue la part de contrôle de l'utilisateur de cette mesure, elles porteront les traces pondérées de deux indices majeurs pour évaluer la qualité implicative.

(Gras et al. 2015)

UN MARIAGE ARRANGE ENTRE L'IMPLICATION ET LA CONFIANCE ?

Régis GRAS¹, Raphaël COUTURIER², Pablo GREGORI³

RÉSUMÉ

La plupart des indices d'association entre variables binaires utilisent la fréquence conditionnelle qu'ils appellent confiance ou expressions algébriques d'instanciations pour décider d'une liaison entre deux variables et en apprécier la qualité. En Analyse Statistique Implicative, une autre mesure, l'intensité d'implication, vise le même objectif en se limitant à l'implication tout en s'appuyant plutôt sur la probabilité d'apparition des contre-exemples à ce type de liaison. Dans cet article nous comparons ces deux mesures en montrant qu'elles ont des propriétés mais possèdent des relations analytiques intéressantes. De ce fait, nous concevons et expérimentons une nouvelle mesure de qualité d'implication en deux approches qui associe confiance et intensité. Nous montrons l'intérêt présenté par cette combinaison pour intégrer la contraposée de l'implication, condition nécessaire pour faire jouer à ce nouvel indice une fonction d'analyse causale.

Mots-clés : confiance, fréquence conditionnelle, intensité d'implication, intensité entropique, implifiance, règle, quasi-règle, contraposée

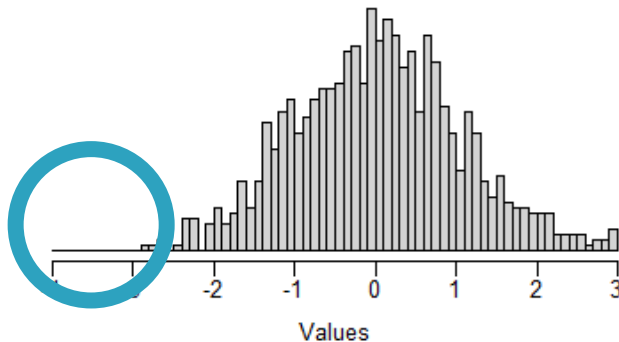
Comportement des indices implicatifs en fonction de la forme des données

Comportement des indices

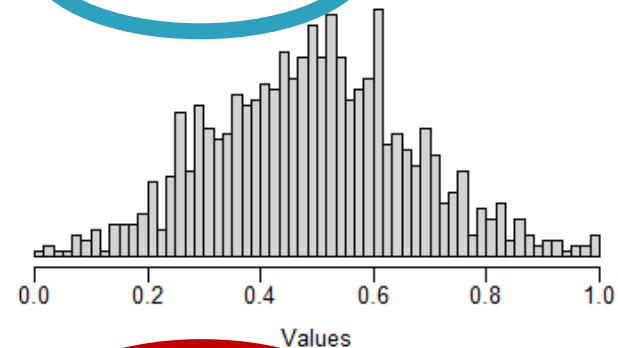
La renormalisation est particulièrement sensible aux outliers

→ On devra s'assurer de l'absence de valeurs extrêmes (winsorization ?)

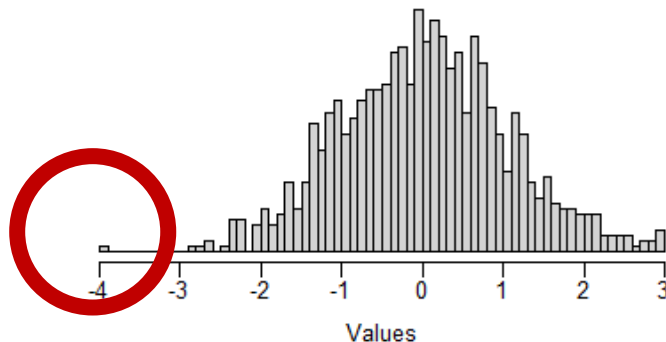
Mean of A = 0.03 | Variance of A = 1.07



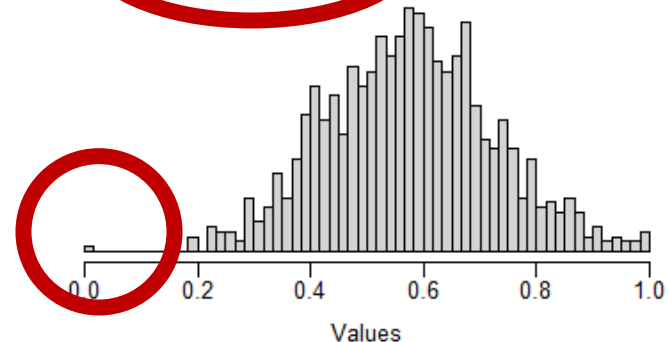
Mean of A = 0.49 | Variance of A = 0.03



Mean of A = 0.02 | Variance of A = 1.08



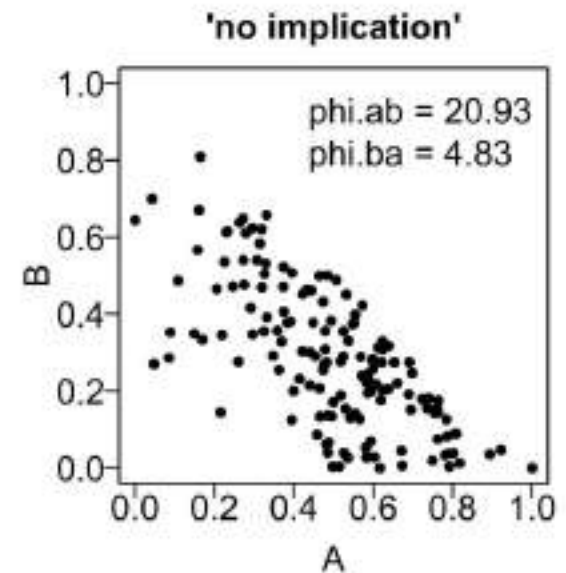
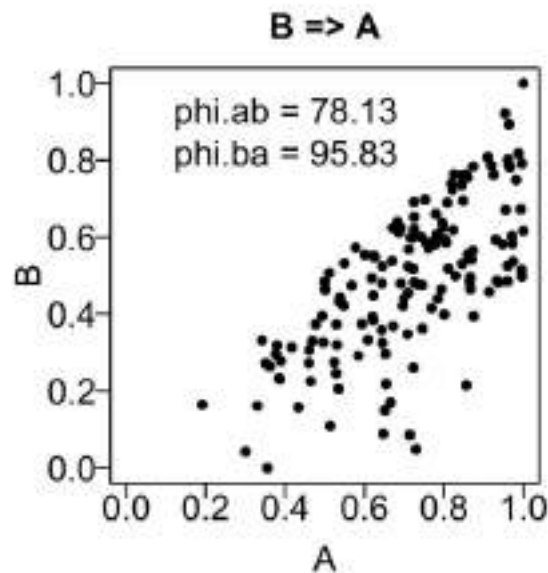
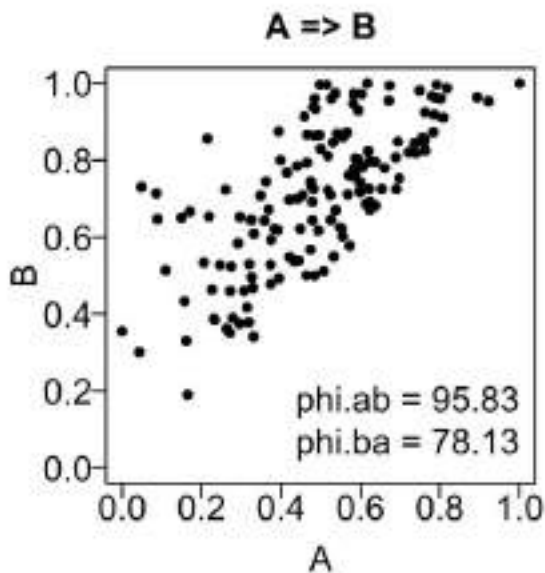
Mean of A = 0.57 | Variance of A = 0.02



Comportement des indices

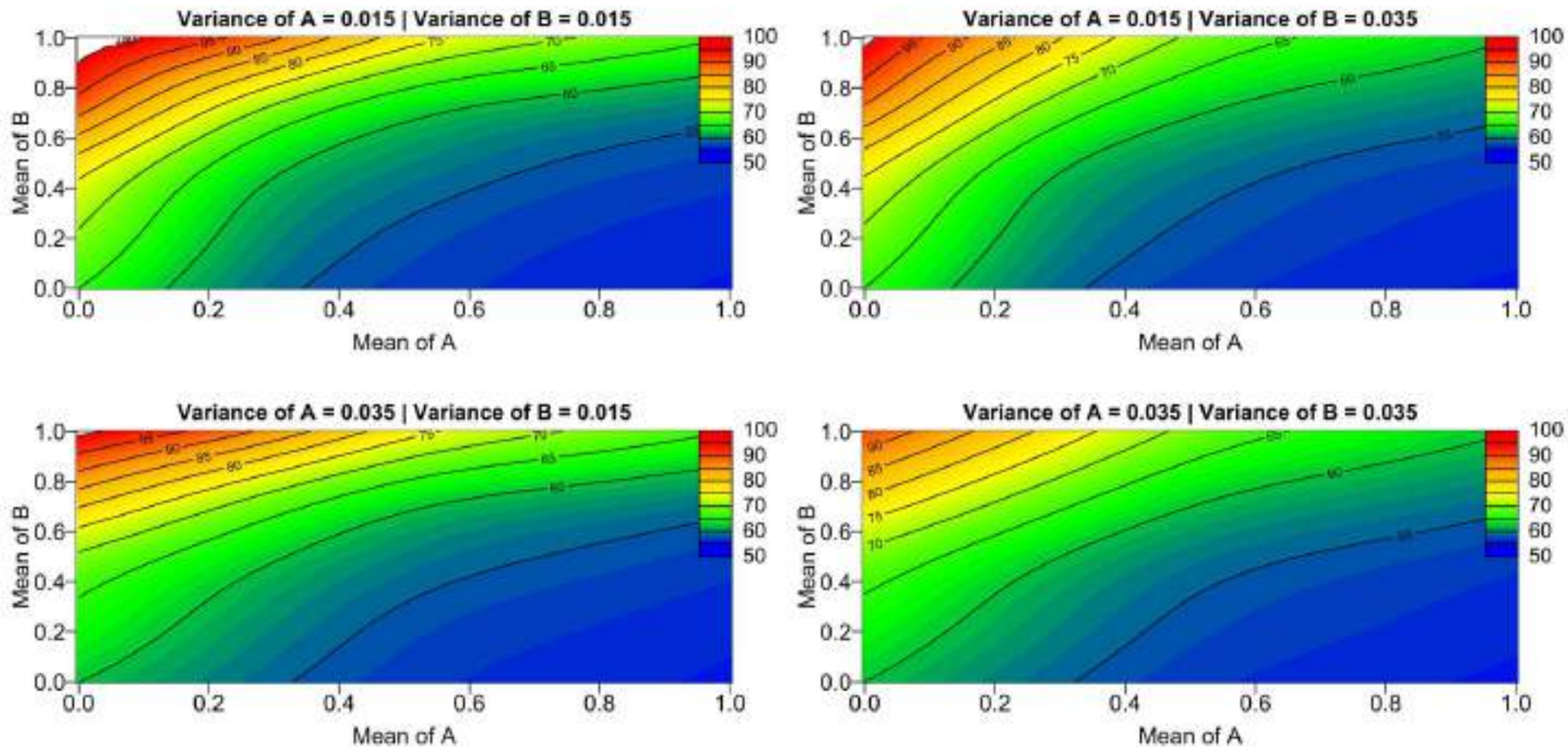
- $r(A,B) > 0$ et $m_A < m_B$: $A \Rightarrow B$
- $r(A,B) > 0$ et $m_A > m_B$: $B \Rightarrow A$
- $r(A,B) < 0$: *rien*

➔ On devra recoder toutes les corrélations sous forme positive...



Comportement des indices

L'indice phi est maximal pour une corrélation parfaite (covariance=variances), $m_A \ll m_B$, et des variances homogènes



Comportement des indices

Le coefficient phi a deux problèmes

- Problème 1 : il fonctionne sous une hypothèse d'indépendance entre les variables
 - Il tend vers 100 lorsque les variables sont corrélées
 - Et des variables avec implication sont forcément corrélées !
- Problème 2 : il dépend fortement de la taille d'échantillon
 - Il tend vers 100 lorsque N augmente

B	0	20
non-B	20	0
	non-A	A

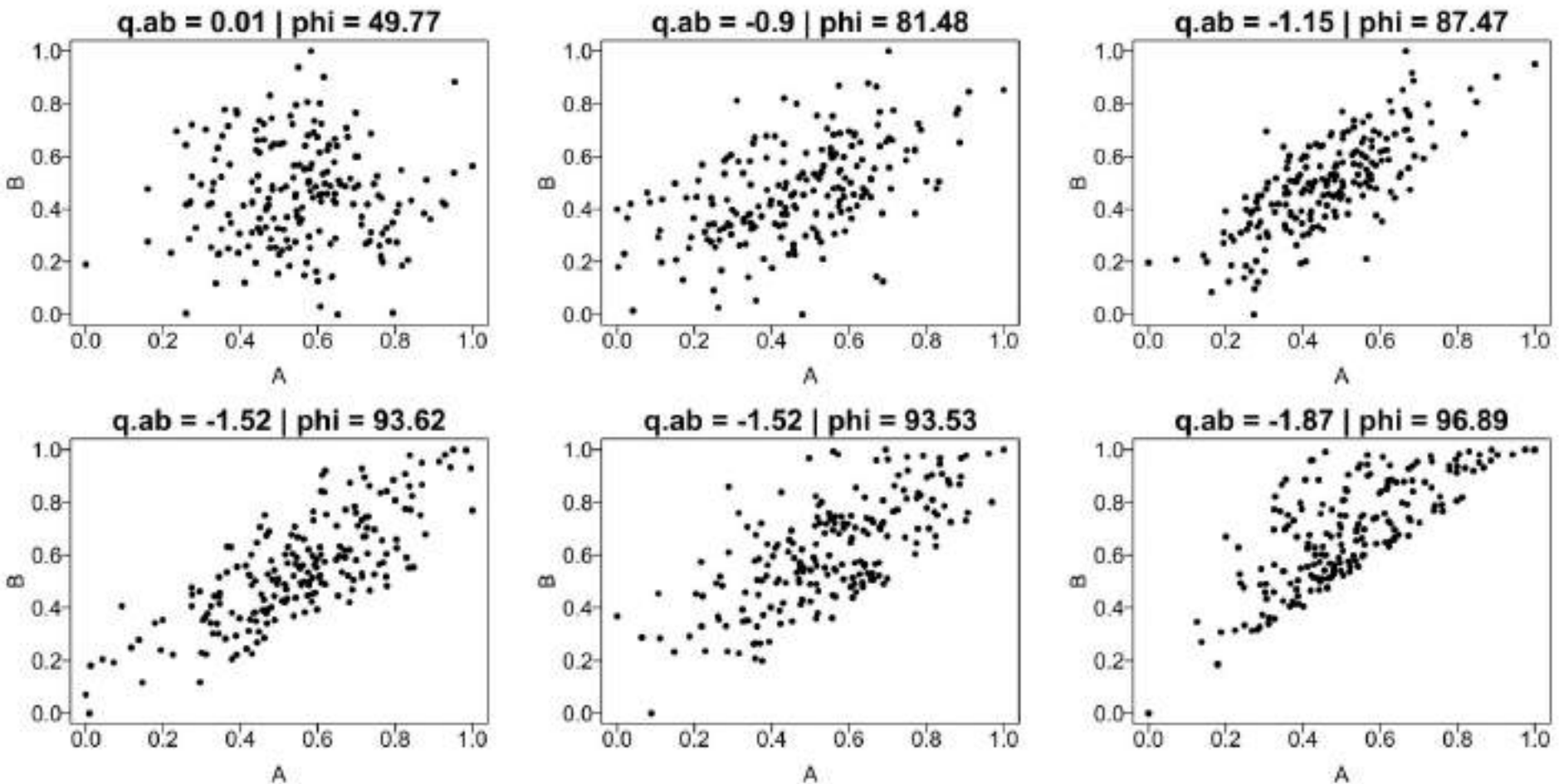
$\text{phi} \approx 100$

B	2000	2000
non-B	2000	1950
	non-A	A

$\text{phi} \approx 100$

Comportement des indices

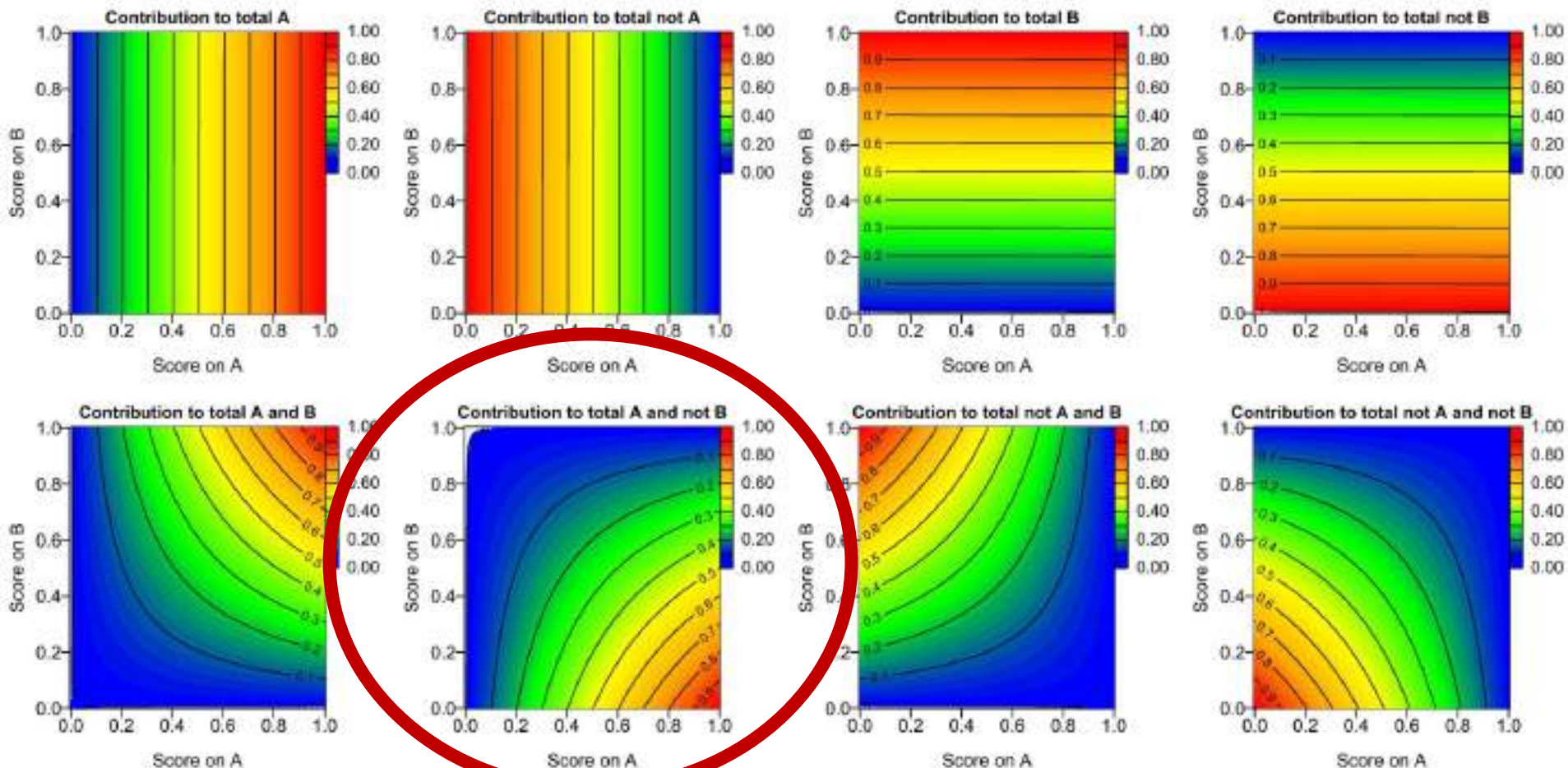
Coefficient phi sur données simulées



Comportement des indices

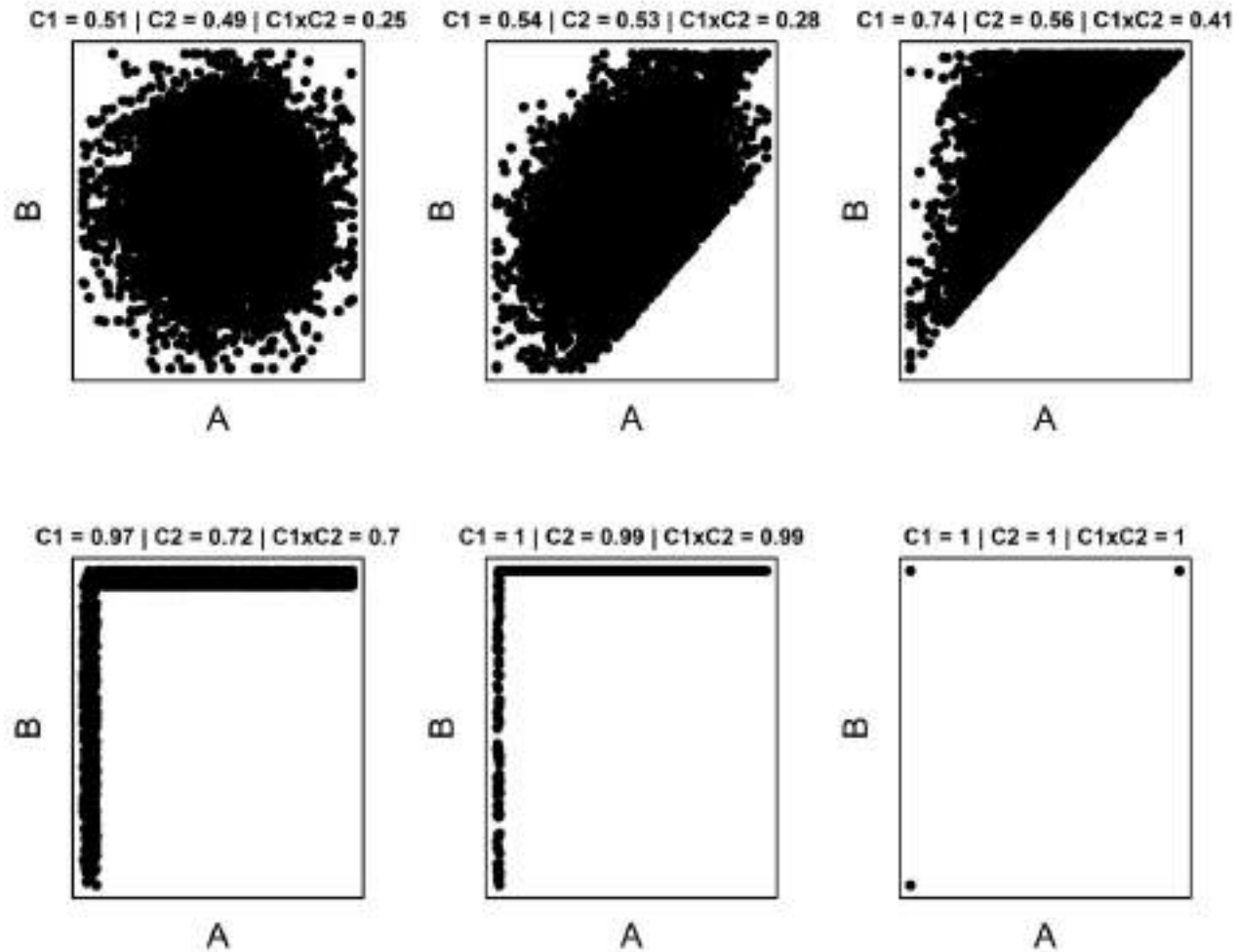
La confiance s'appuie sur la somme des scores "A et non-B". Chaque point contribue en proportion variable... et les seuls points qui ne comptent pas au titre de "A et non-B" sont ceux pour lesquels $A=0$ et/ou $B=1$.

→ La confiance est maximale si les données ont une forme de Γ



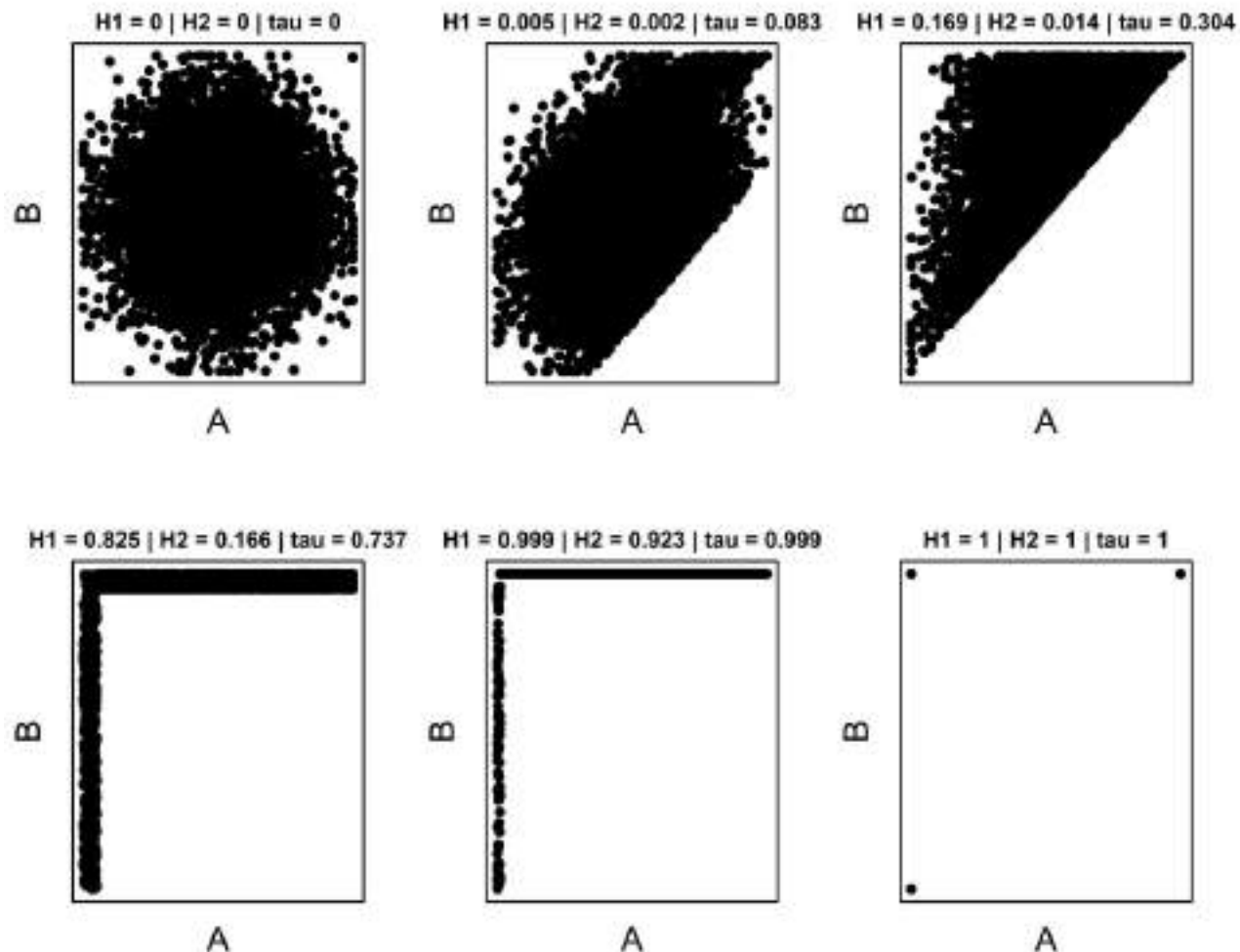
Comportement des indices

Exemple pour la confiance brute



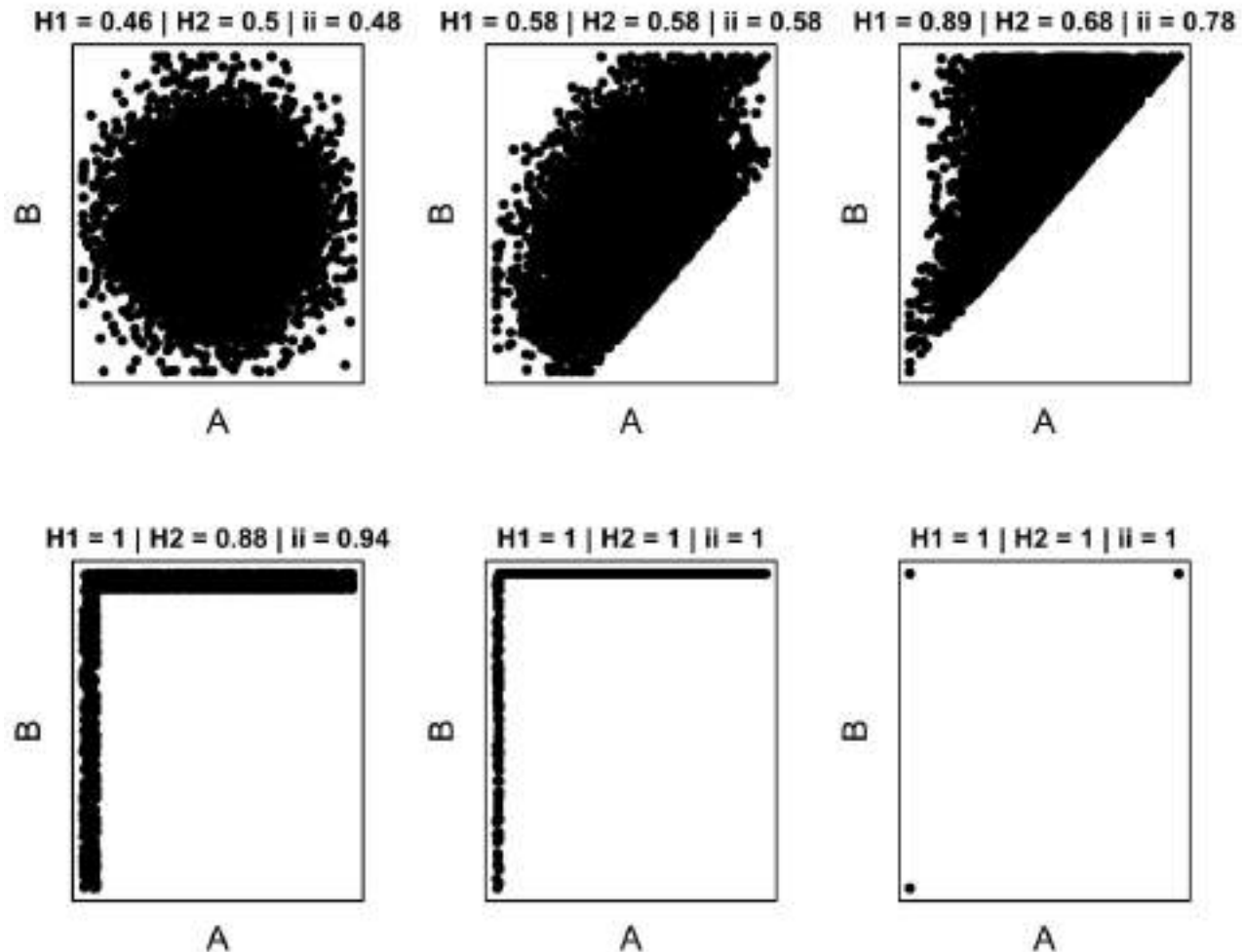
Comportement des indices

Exemple pour la confiance entropique (version 2001)



Comportement des indices

Exemple pour la confiance entropique (version 2012)



Sélection des indices

Sélection des indices

- L'indice q / ϕ ne semble pas optimal
 - Pas de statistique descriptive indépendante de n , vulnérabilité au cas de variables corrélées
 - (malgré la possibilité de calculer un écart entre $q.ab$ et $q.ba$: Bailleux et al. 2019)
- La confiance brute et la confiance entropique V_1 ne semblent pas optimales
- Exigeantes sur le rejet de l'implication, valeurs attendues non plausibles, gamme de variation très limitée sur des données plausibles
- L'implifiance ne semble pas optimale
- Assez arbitraire, combine aussi les défauts des deux approches précédentes
- Pourquoi pas la confiance entropique 2012 / le coefficient ψ ?

Quelle statistique descriptive ?

- Je veux une statistique descriptive, donc indépendante de N
 - Solution logique : retirer le coefficient modérateur par n
- ➔ S'en tenir à l'indice d'inclusion

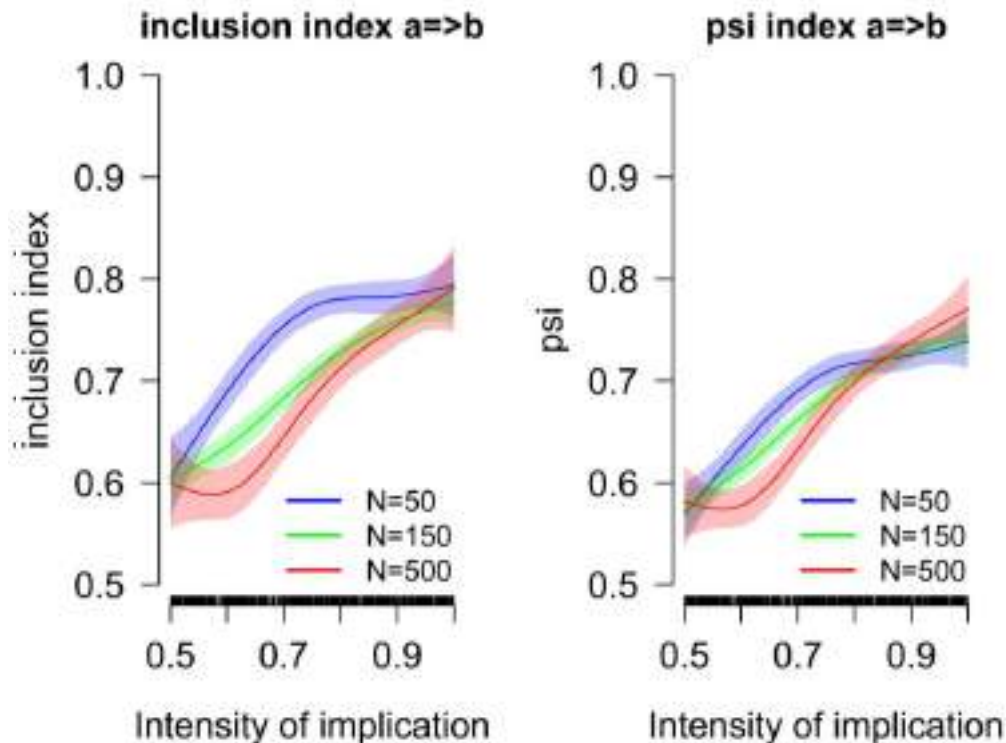
$$i(a, b) = \left(\left[1 - h_1(t) \right] \left[1 - h_2(t) \right] \right)^{\frac{1}{2}}$$

$$\Psi(a, b) = \left(1 - \frac{1}{2\sqrt{n}} \right) i(a, b)$$

Sélection des indices

Quelle statistique descriptive ?

- ...sauf que l'indice d'inclusion varie avec N !
- Entropie plus grande quand N est petit → indice surestimé



- Le coefficient modérateur corrige (partiellement) un biais présent dans l'indice d'inclusion

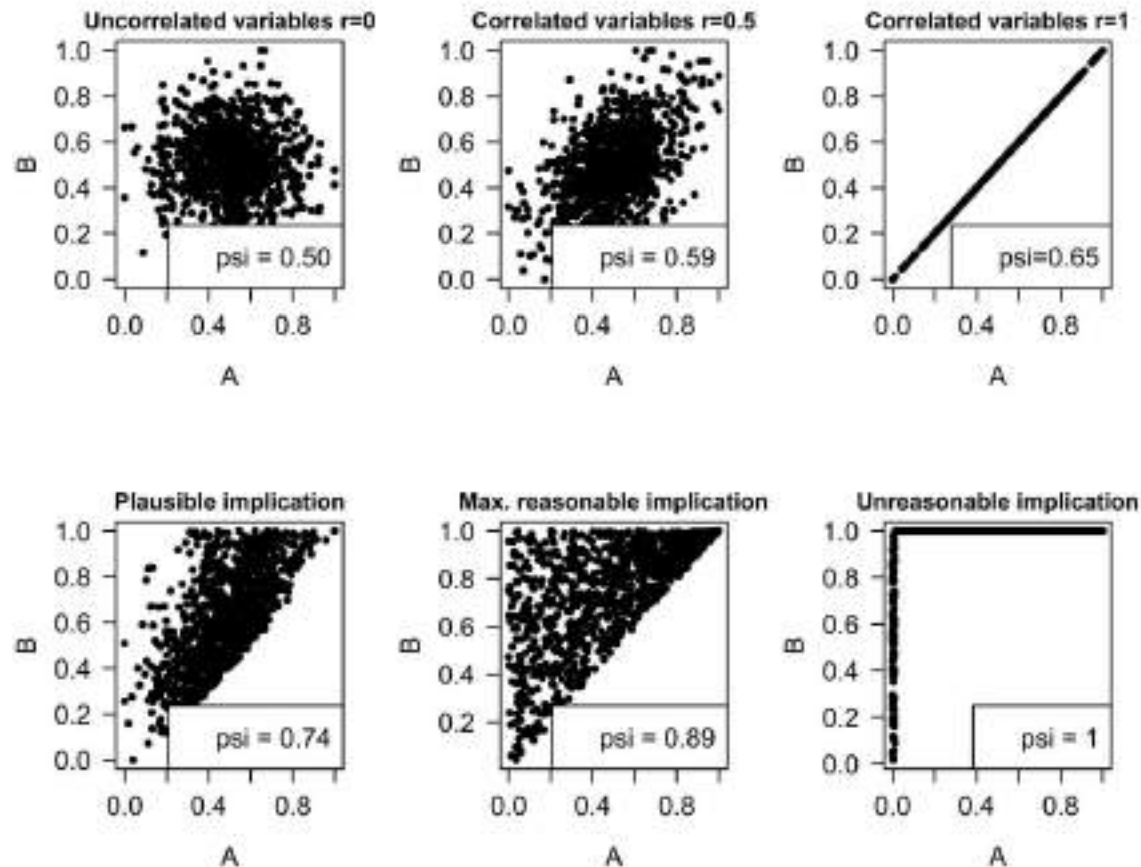
→ On garde psi

Sélection des indices

Quelle statistique descriptive ?

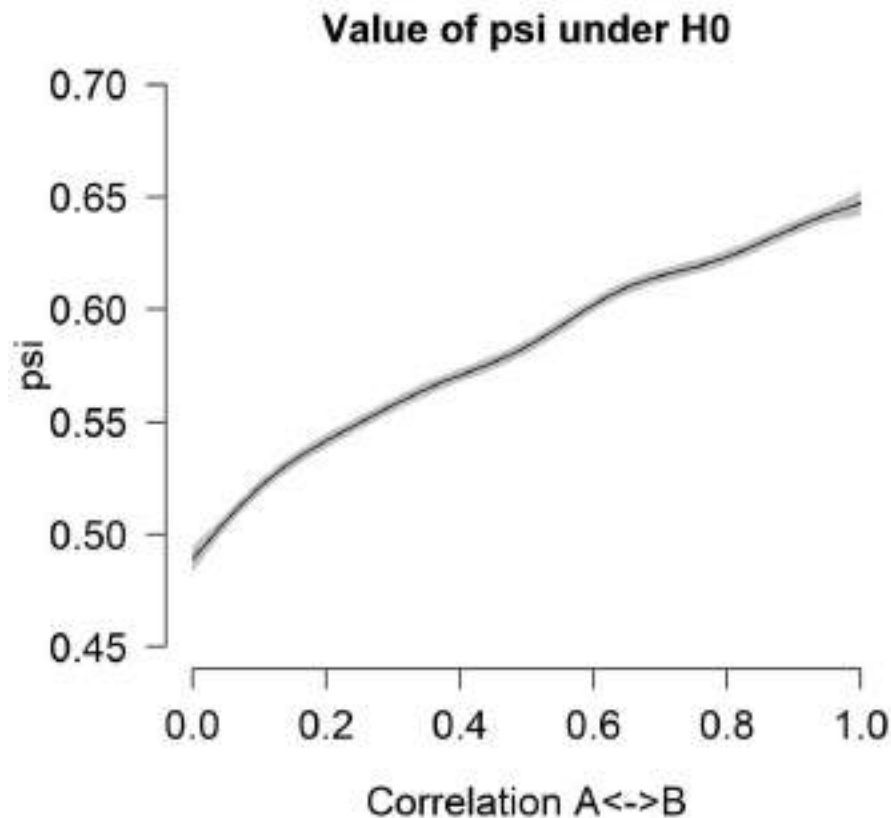
- Le coefficient psi ne varie pas entre 0 et 100
- Il ne vaut pas 0 pour des données indépendantes et n'atteindra jamais 1 sur des données réelles

→ Normaliser par les valeurs minimale et maximale plausibles ?



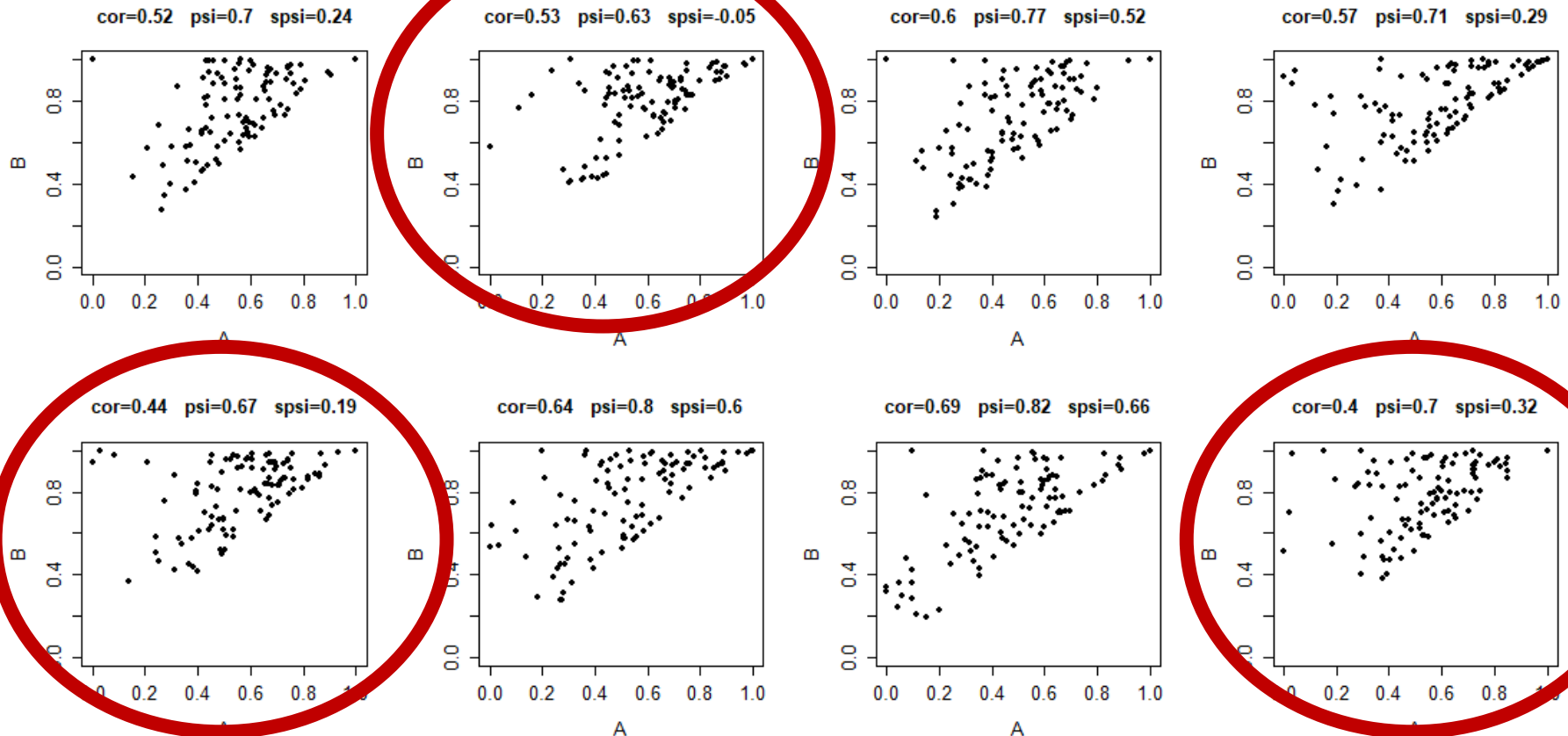
Sélection des indices

- Problème : la valeur minimale dépend de $r(A,B)$
- Calculer la valeur minimale attendue pour une corrélation donnée, à chaque analyse, et en tenant compte de N ?



Sélection des indices

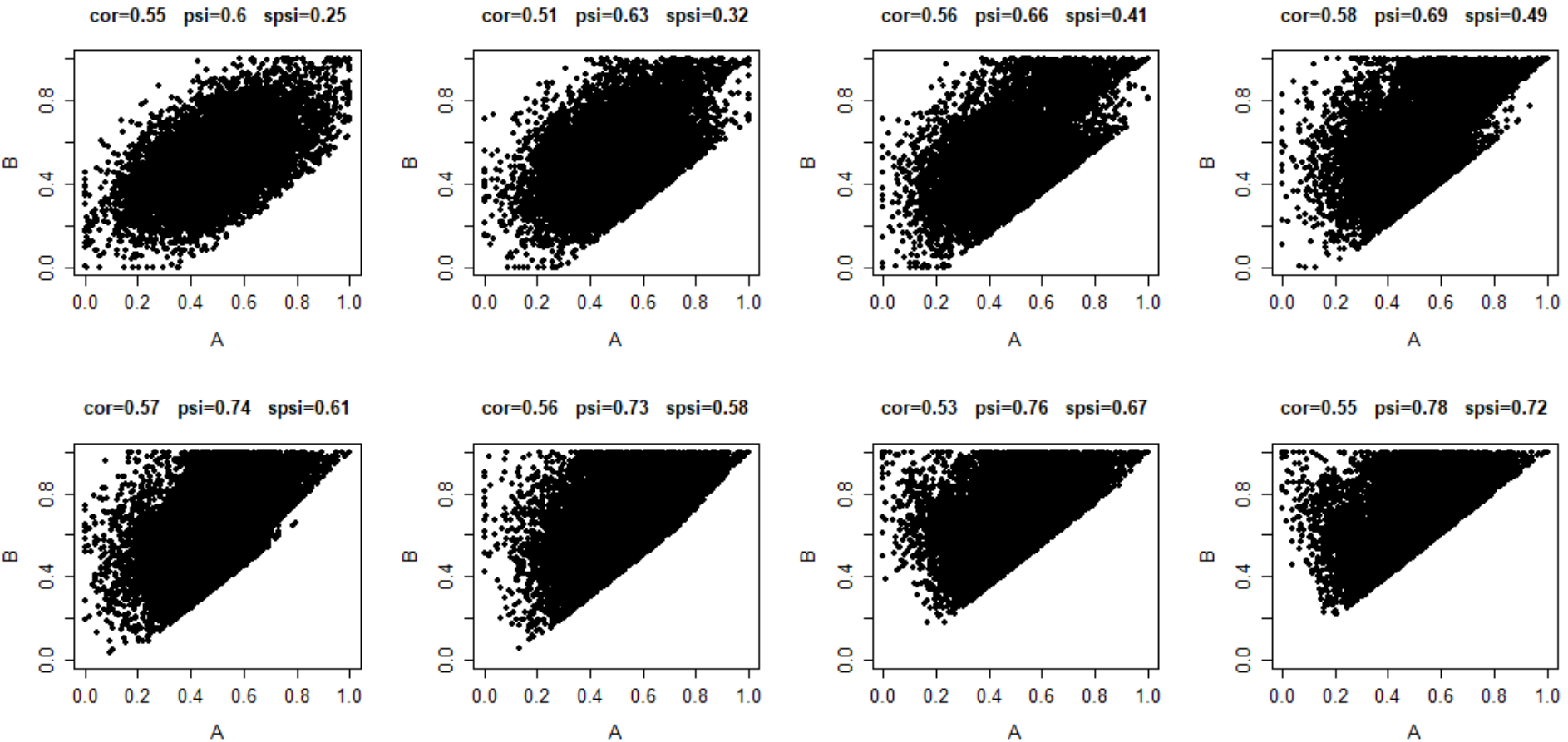
- Problème 2 : normaliser le coefficient par rapport à $r(A,B)$ rend le résultat assez instable
- Huit nuages de points générés avec $n=100$, implication=1 :



Sélection des indices

Meilleure solution (?) : normaliser sur [0.50, 0.89]

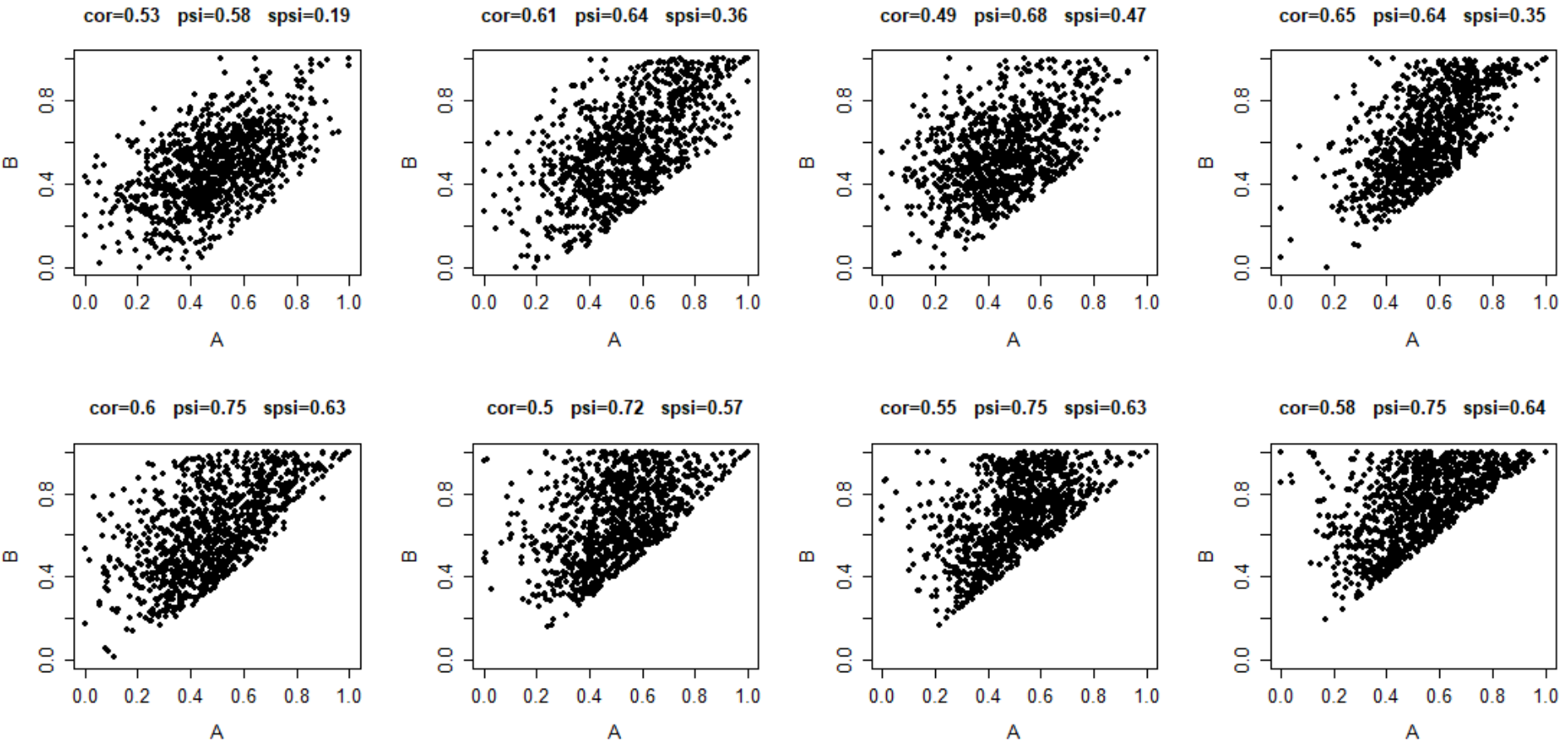
$$N = 10000$$



Sélection des indices

Meilleure solution (?) : normaliser sur [0.50, 0.89]

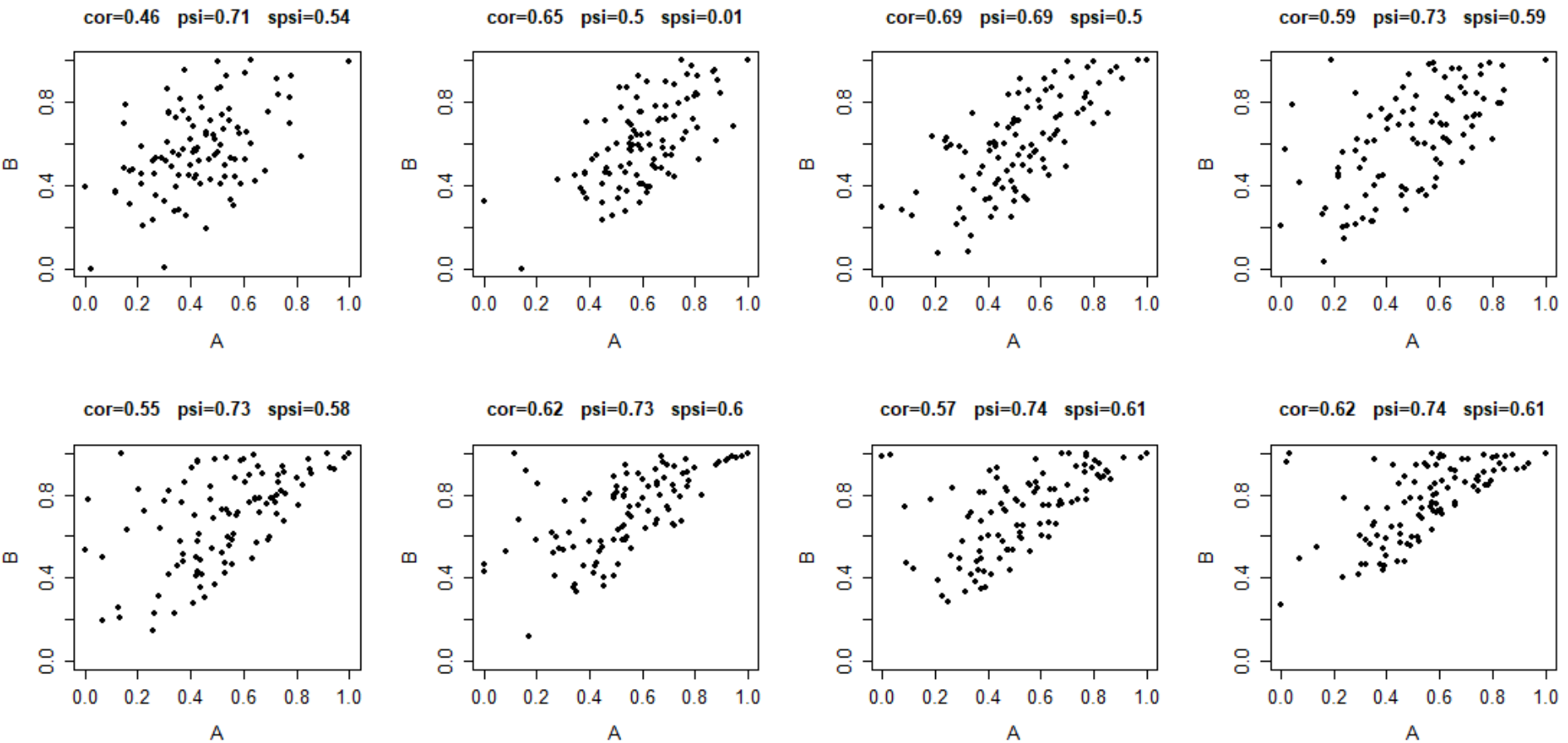
$N = 1000$



Sélection des indices

Meilleure solution (?) : normaliser sur [0.50, 0.89]

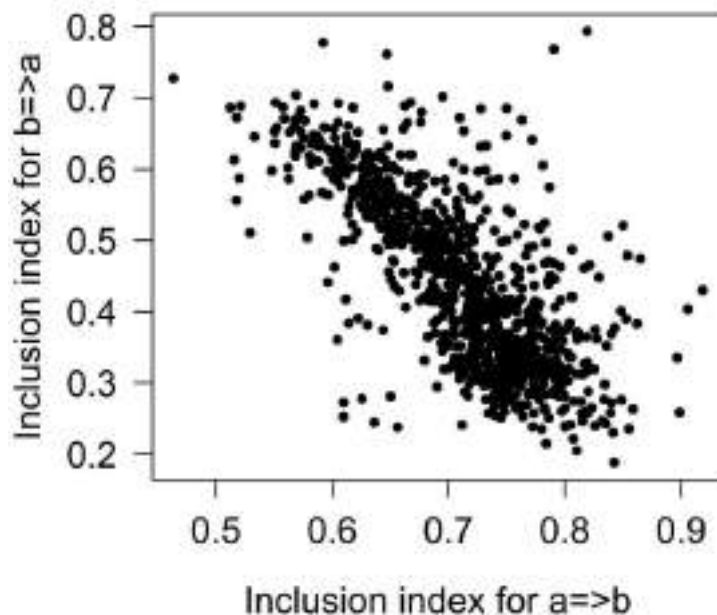
$$N = 100$$



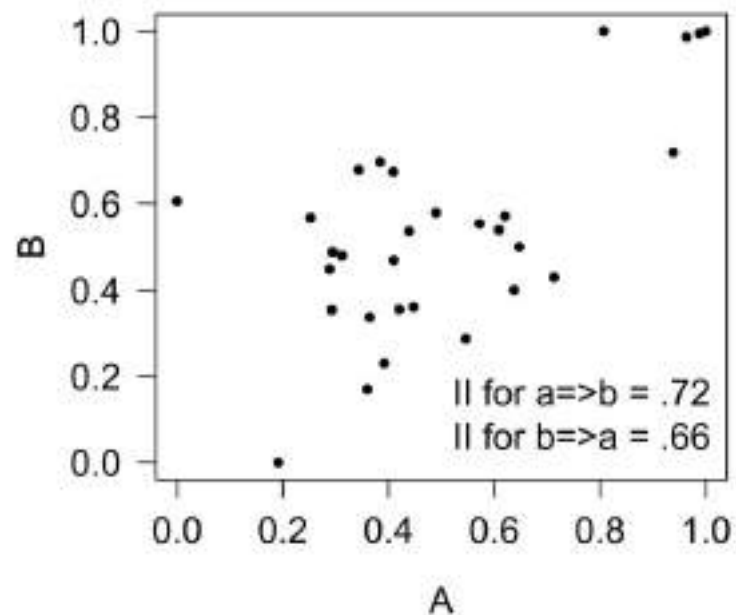
Sélection des indices

- Remarque : pas de relation parfaite entre l'indice pour $a \Rightarrow b$ et pour $b \Rightarrow a$
- ➔ pas possible de normaliser le coefficient autour de 0.5 et de lire un coefficient négatif comme une implication $b \Rightarrow a$ ➔ nécessité de calculer l'indice dans les deux sens ➔ valeurs négatives fixées à 0

Implication in a series of random samples



Example dataset



Sélection des indices

- Quelle statistique inférentielle ?
- Une solution simple et élégante : l'approche Monte Carlo
 - Générer des données sous H_0 avec corrélation connue
 - Normaliser ces données entre 0 et 1 avec winsorization (sinon les indices implicatifs sont plus dispersés quand N augmente !)
 - Calculer l'indice implicatif qui nous intéresse
 - Recommencer (x 10.000)
 - Comparer l'indice réellement obtenu à cette distribution

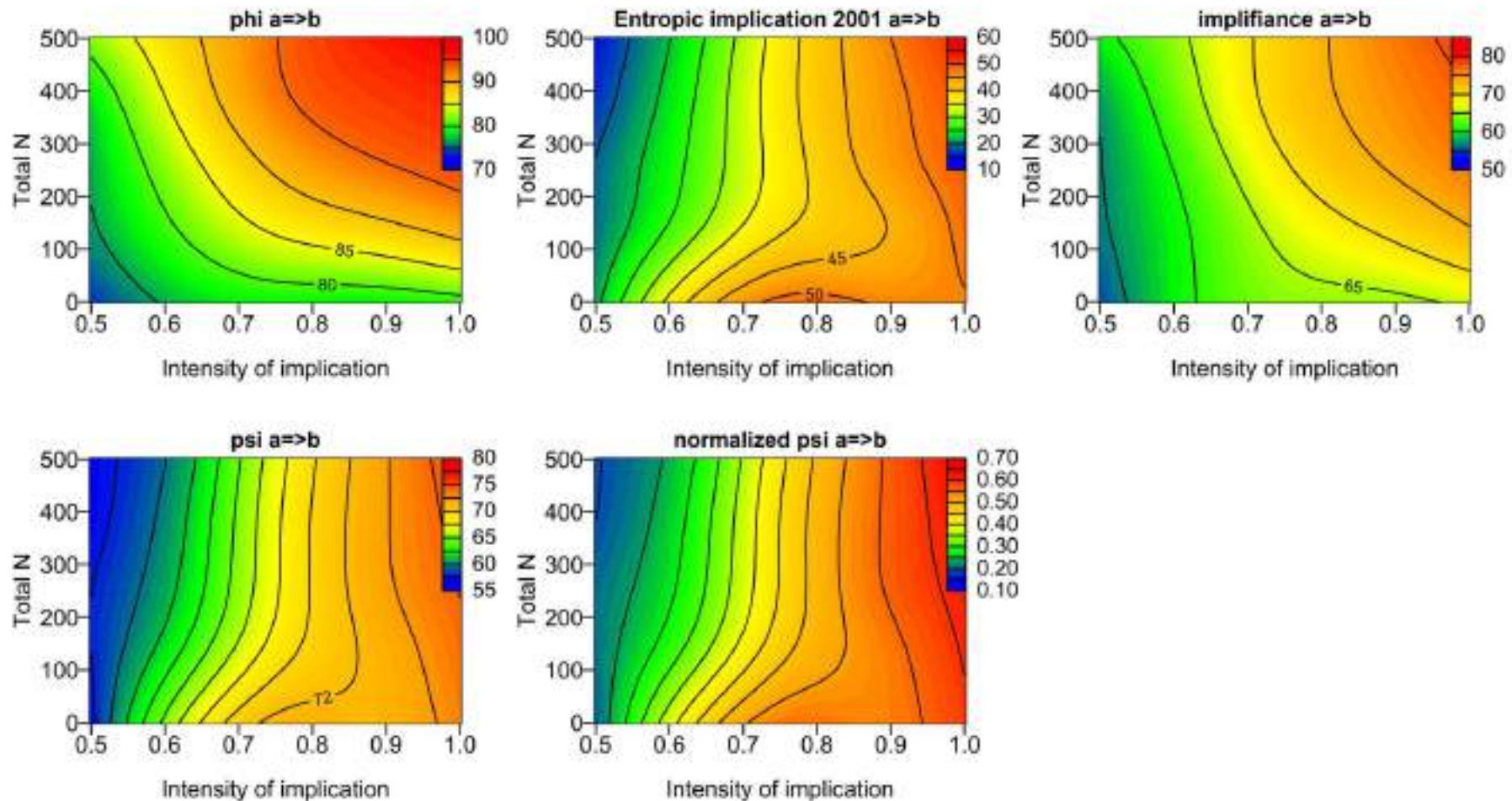
Sélection des indices

1. Winsorizer les données → le placement de la moyenne dépendra surtout de l'asymétrie plutôt que des outliers
2. Normaliser les données → exprimées entre 0 et 1
3. Recoder les corrélations négatives en corrélations positives → pas de risque de passer à côté d'une implication négative
4. Calculer la confiance entropique de Gras & Couturier (2012), sans le coefficient modérateur pour N → EII
5. Standardiser cette confiance entropique entre les bornes $[ii.cor(A,B), 0.89]$ → sEII
6. Calculer la distribution des valeurs de sEII sous H_0 → $p(sEII | H_0)$

Simulations numériques à grande échelle

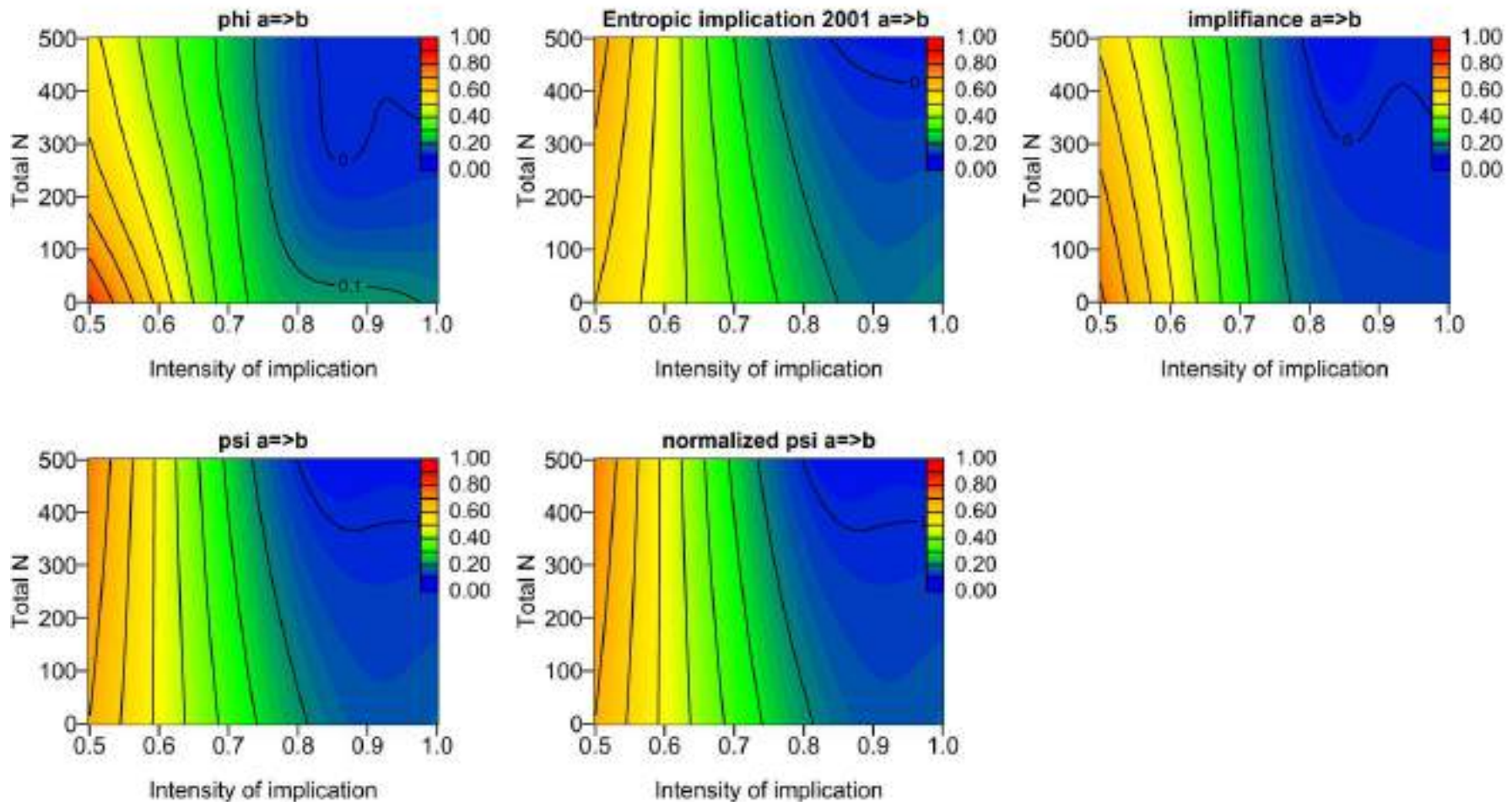
Simulations numériques

Indices d'implication : une sensibilité à N visible pour ϕ



Simulations numériques

Valeurs p : de bons résultats pour toutes les options !

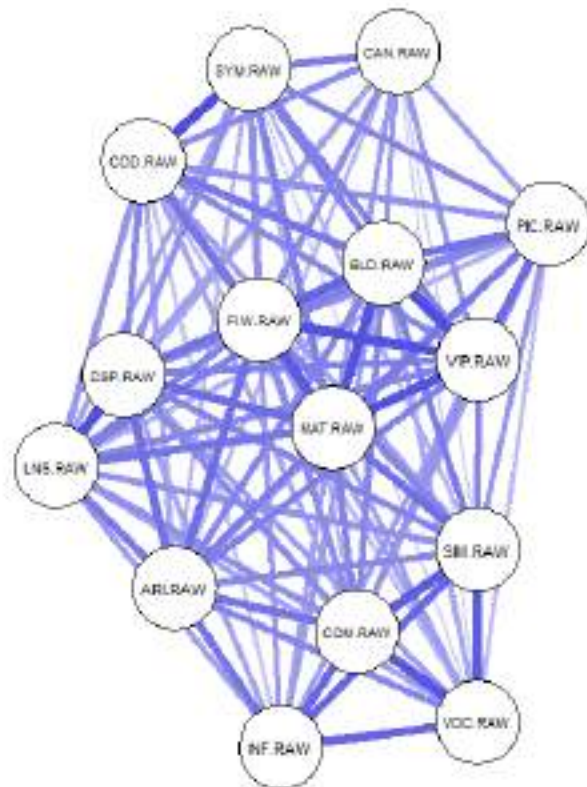


Analyse en réseau implicative

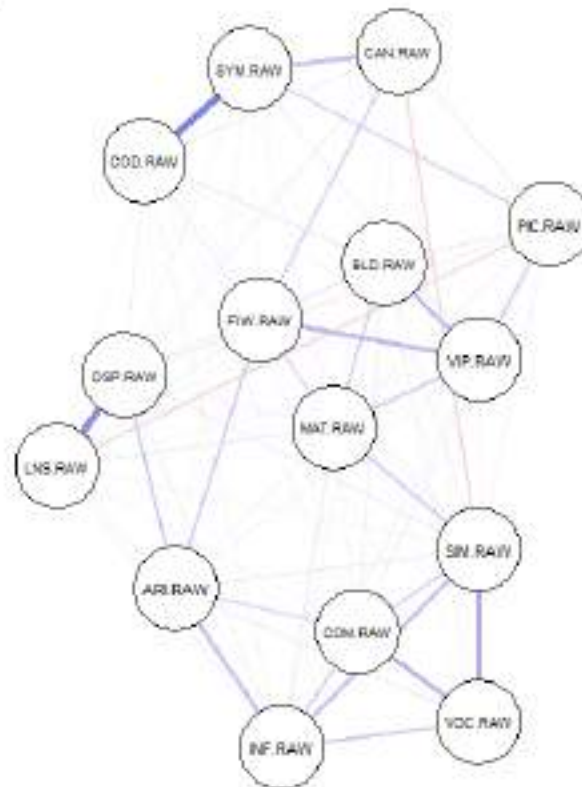
Réseau implicatif

Et si on veut faire un réseau ?

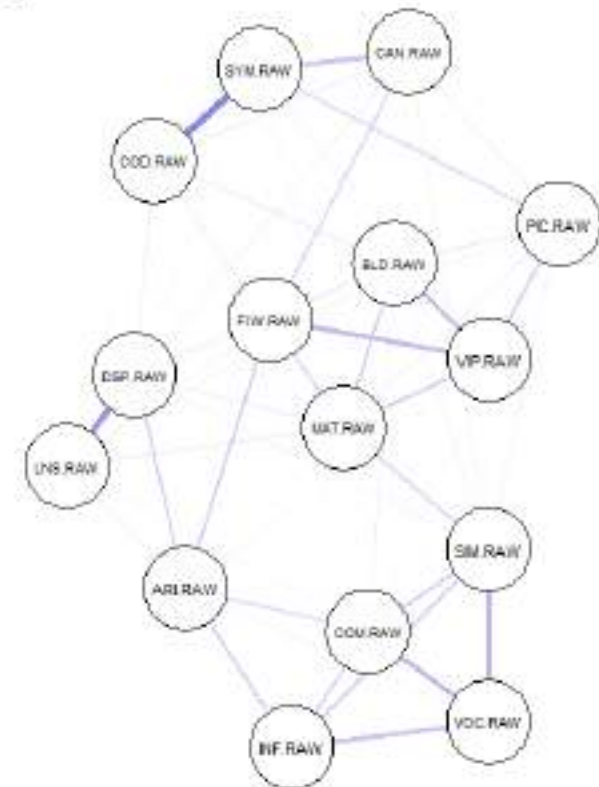
correlations



partial correlations

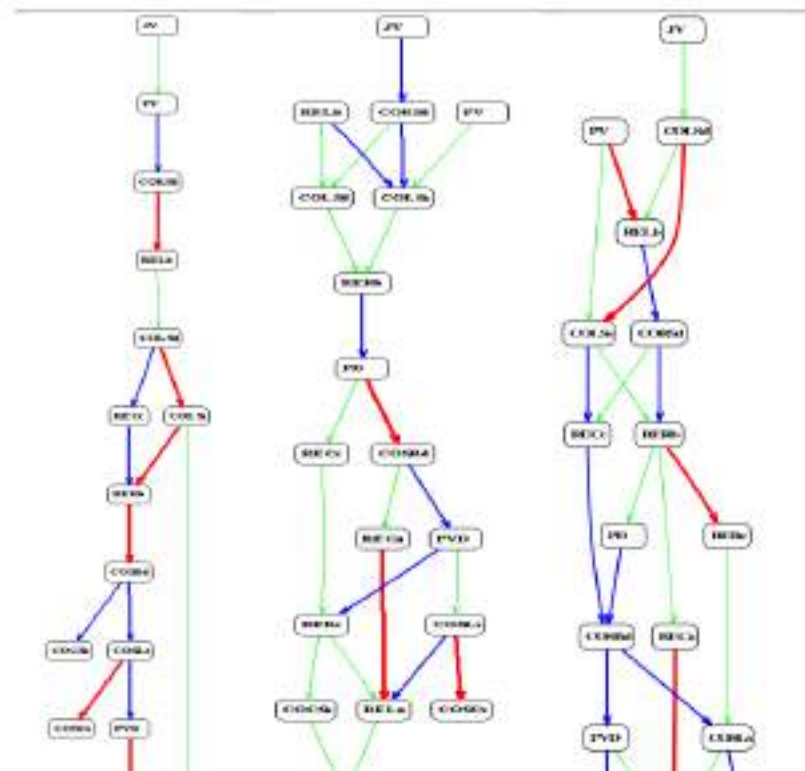


glasso EBIC



Réseau implicatif

- Des options déroutantes : arbres implicatifs cohésitifs et méta-règles de Gras et al.



Réseau implicatif

Des options reposantes : la méthode Dauvier

1. On représente l'implication (plutôt que r_{ab} ou $r_{ab.c}$)
2. Régularisation par valeurs p (plutôt que lasso)
3. Réseau dirigé (plutôt que bidirectionnel)
4. Et roulez jeunesse



Exemples d'application



<https://osf.io/c5h78/>

```
#####
# Examples
#####

# Generate a dataset containing two variables A and B with an implicative shape
testDataset <- implicationSimulatedData(n=200, corAB=0.50, implicationThreshold=0.90, winsorize=0.001)

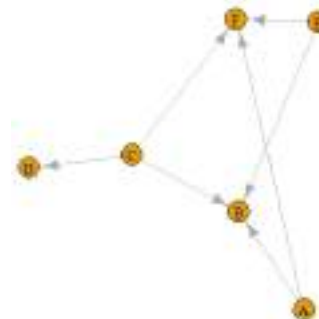
# Compute npsi index of implication for both a=>b and b=>a, as well as the corresponding p-values with 1000 samples
# Notes: 10.000 samples is preferable for actual use ; normalize is set to FALSE because the dataset generated above is already normalized and winsorize
npsi <- ASI.npsi(testDataset$A.norm, testDataset$B.norm, normalize=F, winsorize=0.001, p=TRUE, samples=1000)

# Plot A and B and report the implication values and p-values
plot(testDataset$A.norm, testDataset$B.norm, pch=20, xlab="A", ylab="B")
legend(x="bottomright", legend=c(paste("npsi.a=>b = ", round(npsi$npsi.ab, 3), ", p=", round(npsi$npsi.ab.p, 3), sep=""), paste("npsi.b=>a = ", round(npsi$npsi.ba, 3), ", p=", round(npsi$npsi.ba.p, 3), sep="")), bty="n")

# Generate a longer dataset containing six variables with implicative relationships
testDataset1 <- implicationSimulatedData(n=200, corAB=0.50, implicationThreshold=0.90, winsorize=0.001)[,2:3]
testDataset2 <- implicationSimulatedData(n=200, corAB=0.40, implicationThreshold=0.80, winsorize=0.001)[,2:3]
testDataset3 <- implicationSimulatedData(n=200, corAB=0.30, implicationThreshold=0.70, winsorize=0.001)[,2:3]
fullDataset <- data.frame(cbind(testDataset1, testDataset2, testDataset3))
colnames(fullDataset) <- c("A","B","C","D","E","F")

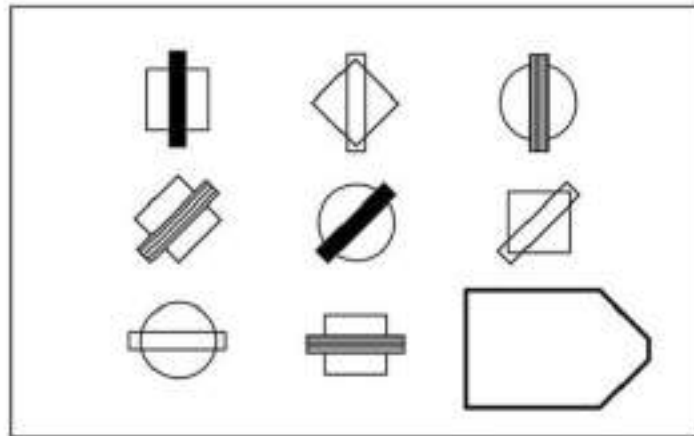
# Draw a directed implicative network, showing only relations that are significant at p<.05
# Notes: 10.000 samples is preferable for actual use ; normalize is set to FALSE because the dataset generated above is already normalized and winsorize
# Can take a minute to run
generateImplicationGraph(dataset=fullDataset, firstColumn=1, normalize=F, winsorize=0.001, samples=1000, threshold=0.05)
```

	A	B	C	D	E	F
A	0	0.659	0	0.000	0	0.224
B	0	0.000	0	0.000	0	0.000
C	0	0.567	0	0.582	0	0.308
D	0	0.000	0	0.000	0	0.000
E	0	0.459	0	0.000	0	0.403
F	0	0.000	0	0.000	0	0.000



Exemples d'application

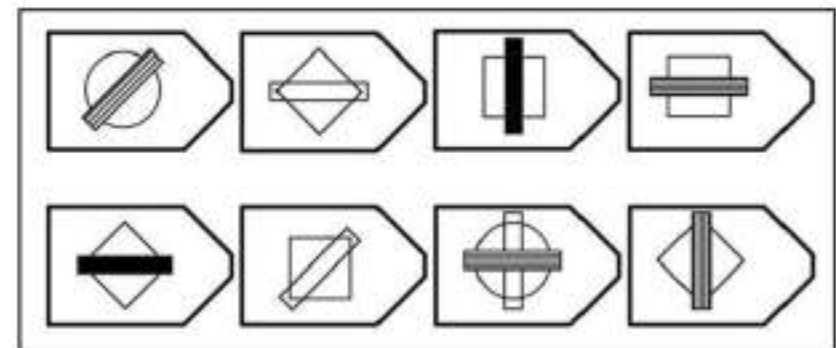
Matrices de Raven



a



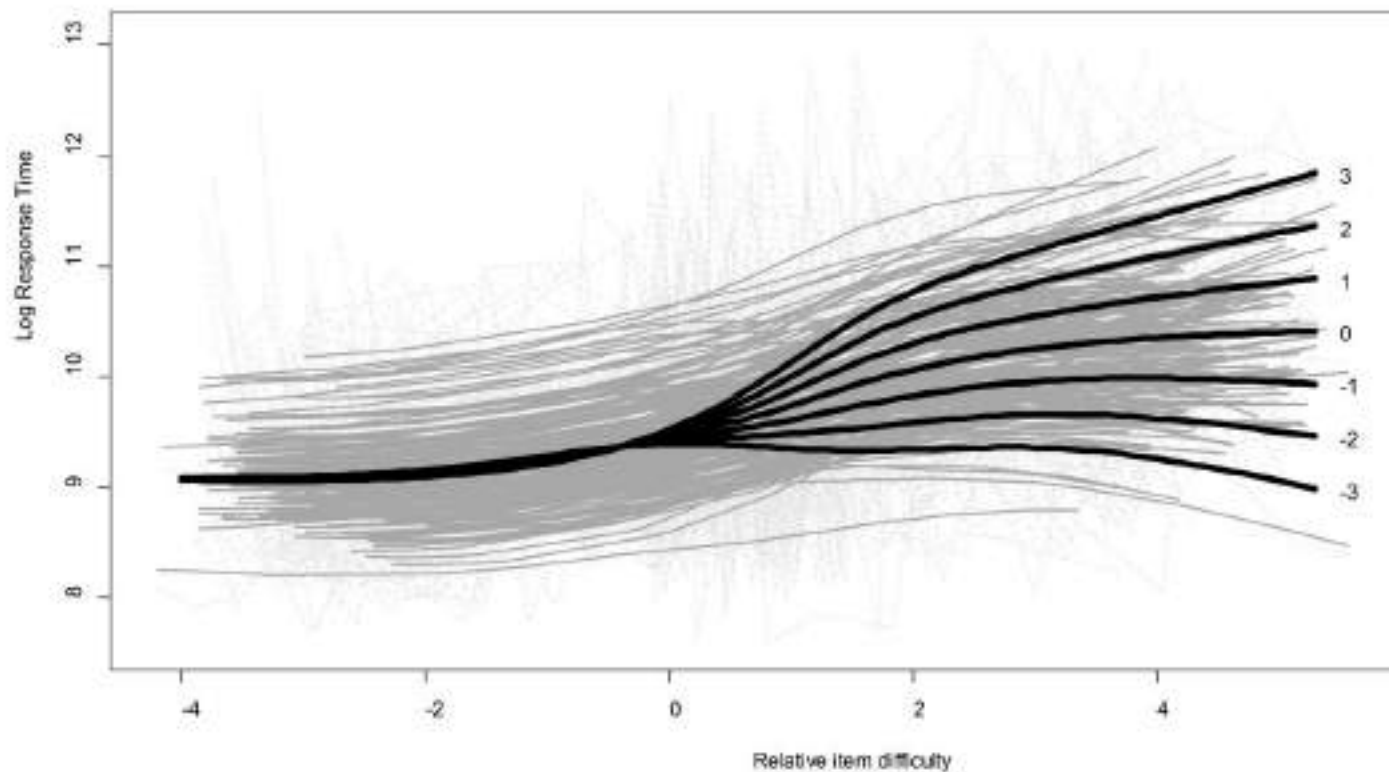
b



- Score total
- Stratégie efficace : proportion de temps sur la matrice

Exemples d'application

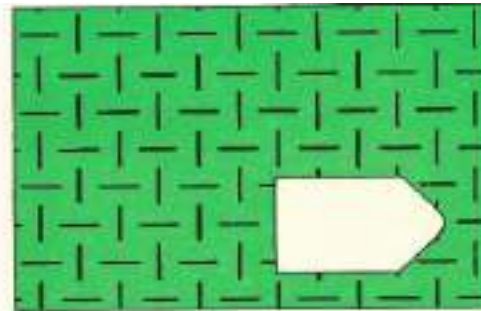
Matrices de Raven



- Temps total
- Niveau de modulation : $r(\text{Difficulté}, \text{TR})$

Exemples d'application

Matrices de Raven



Regarde
bien le
dessin...

...et clique où tu veux sur cette ligne,
pour nous dire si tu penses qu'il est dur.



- Niveau d'estimation métacognitive : $r(\text{Jugement}, \text{Difficulté})$
- Niveau de calibration métacognitive : $r(\text{Confiance}, \text{Réussite})$

Exemples d'application

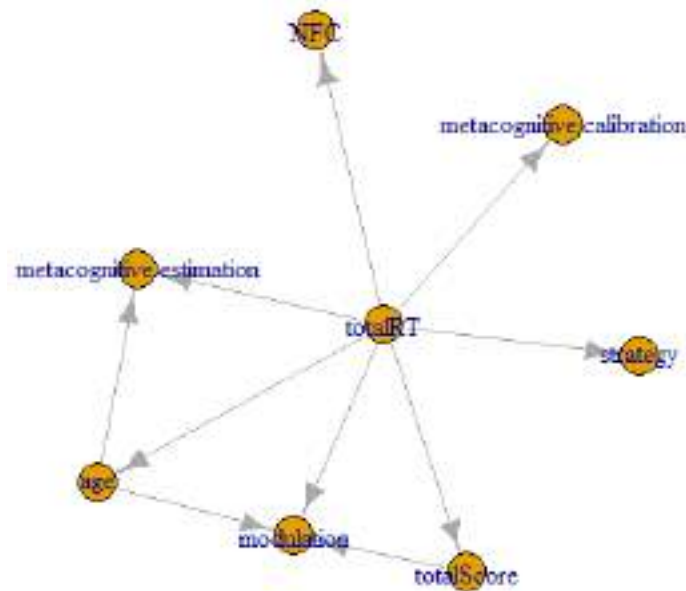
Matrices de Raven

procedure	NFCID	NFCScale	NFCItem
NFC	01	S	J'adore réfléchir pour trouver la solution à un problème
NFC	02	R	Je trouve que réfléchir très fort et longtemps est un peu ennuyeux
NFC	03	S	Je préfère les problèmes compliqués aux problèmes simples
NFC	04	R	Réfléchir ne m'amuse pas vraiment
NFC	05	S	Je suis quelqu'un qui aime beaucoup réfléchir
NFC	06	R	Je préfère faire quelque chose où il ne faut pas beaucoup réfléchir plutôt que quelque chose où il faut beaucoup réfléchir
NFC	07	S	Je suis content quand j'ai fini un travail qui me demandait de beaucoup réfléchir
NFC	08	R	Je réfléchis seulement quand j'y suis obligé

- Besoin de cognition
- Âge de l'enfant

Exemples d'application

Données RavenMeta1 ($N = 174$)

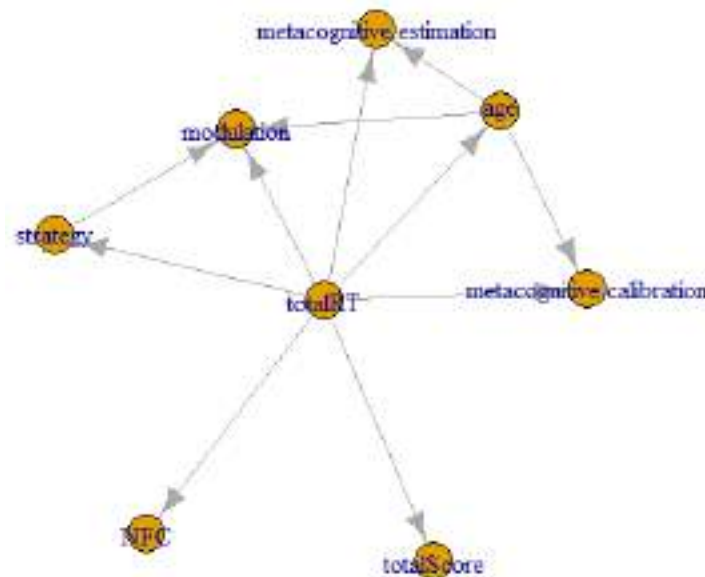


```
> generateImplicationGraph(RM1,1,0.001,100)
```

	totalRT	NFC	metacognitive.estimation	totalScore	metacognitive.calibration	modulation	strategy	age
totalRT	0	0.592	0.768	0.713	0.748	0.863	0.833	0.353
NFC	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
metacognitive.estimation	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalScore	0	0.000	0.000	0.000	0.000	0.544	0.000	0.000
metacognitive.calibration	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
modulation	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
strategy	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
age	0	0.000	0.291	0.000	0.000	0.165	0.000	0.000

Exemples d'application

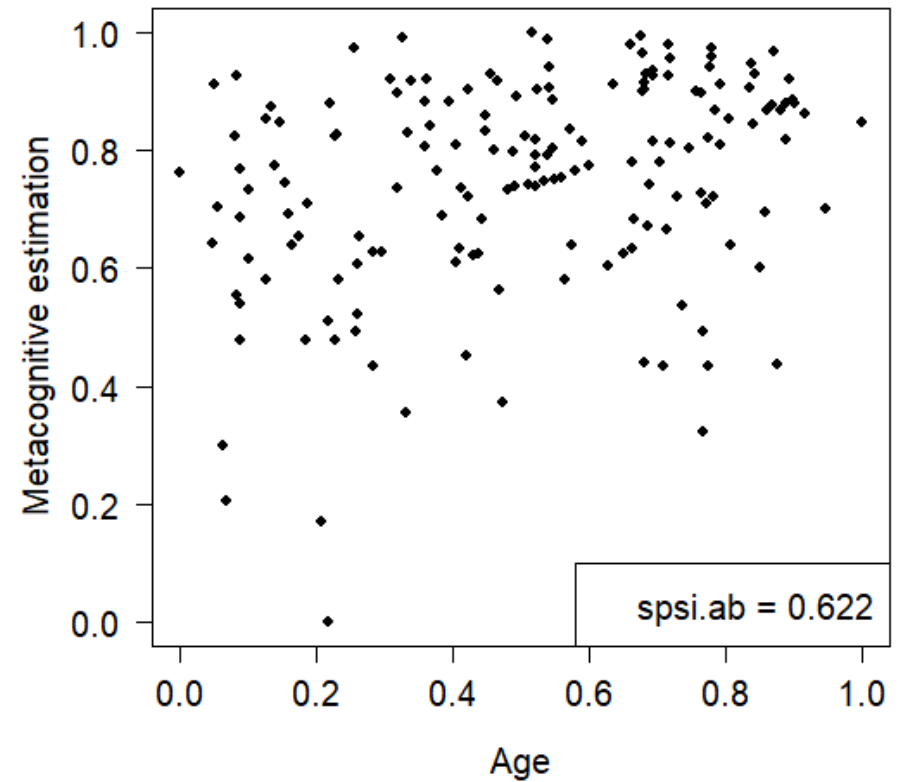
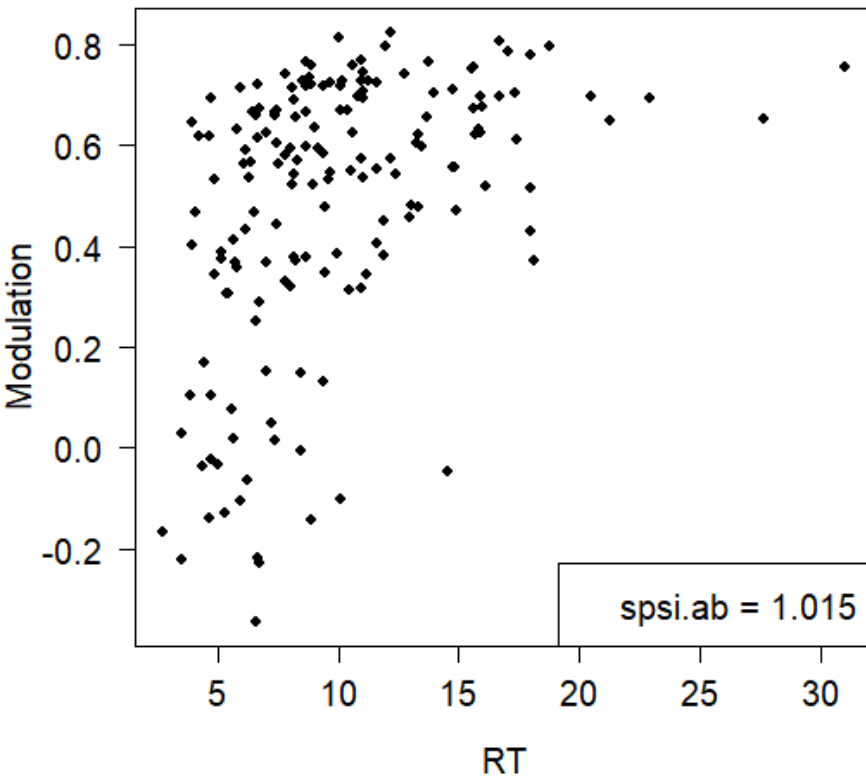
Données RavenMeta2 ($N = 165$)



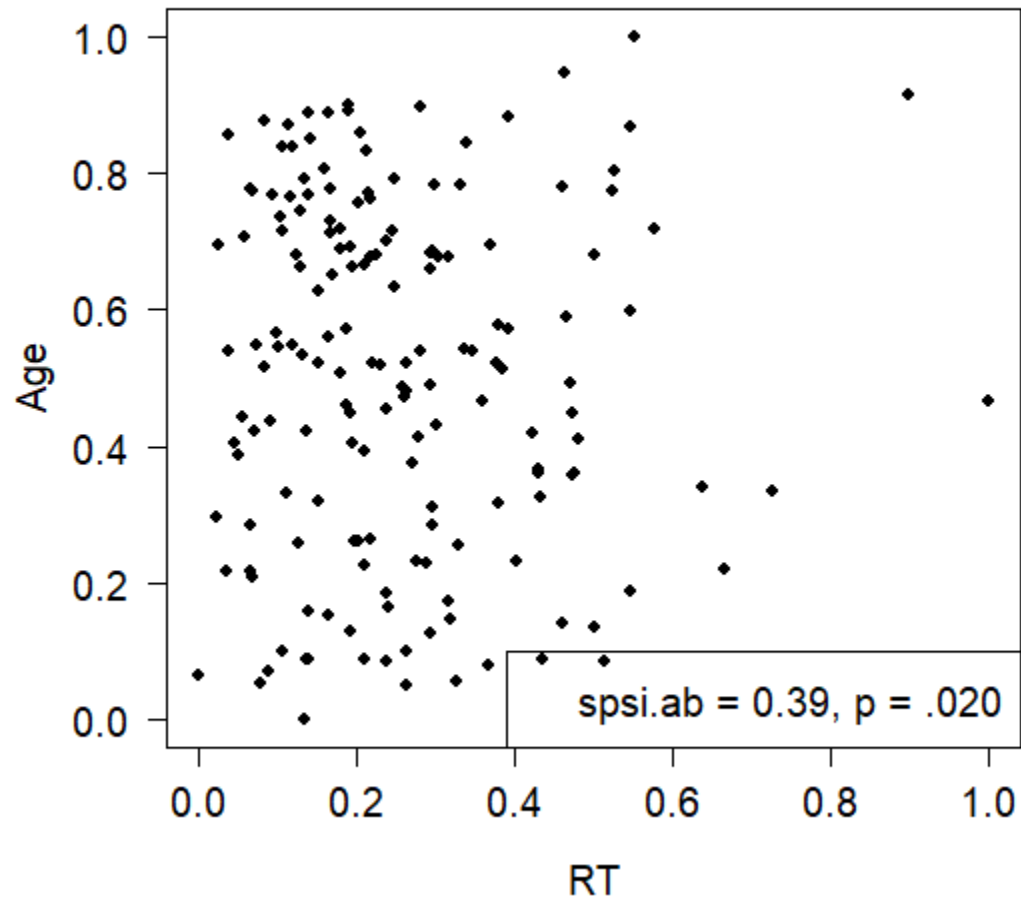
```
> generateImplicationGraph(RM2,1,0.001,100)
```

	totalRT	NFC	metacognitive.estimation	totalScore	metacognitive.calibration	modulation	strategy	age
totalRT	0	0.782	0.997	0.824	0.913	1.000	0.91	0.39
NFC	0	0.000	0.000	0.000	0.000	0.000	0.00	0.00
metacognitive.estimation	0	0.000	0.000	0.000	0.000	0.000	0.00	0.00
totalScore	0	0.000	0.000	0.000	0.000	0.000	0.00	0.00
metacognitive.calibration	0	0.000	0.000	0.000	0.000	0.000	0.00	0.00
modulation	0	0.000	0.000	0.000	0.000	0.000	0.00	0.00
strategy	0	0.000	0.000	0.000	0.000	0.459	0.00	0.00
age	0	0.000	0.622	0.000	0.620	0.644	0.00	0.00

Exemples d'application



Exemples d'application

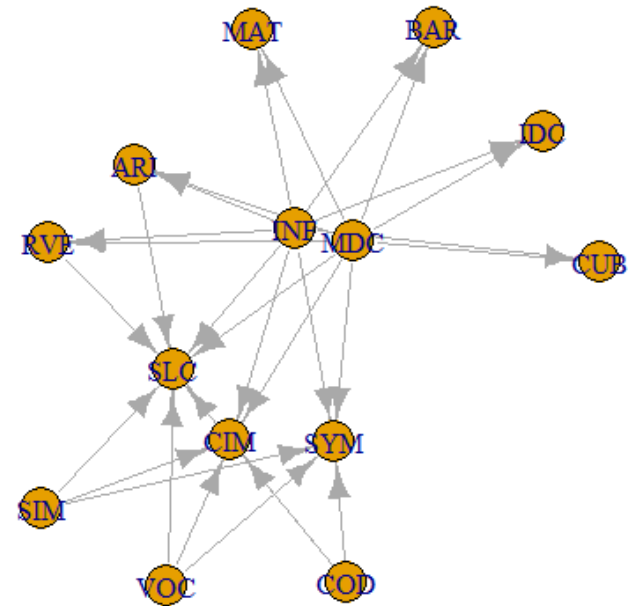


Exemples d'application

Données WISC4 (Lecerf)

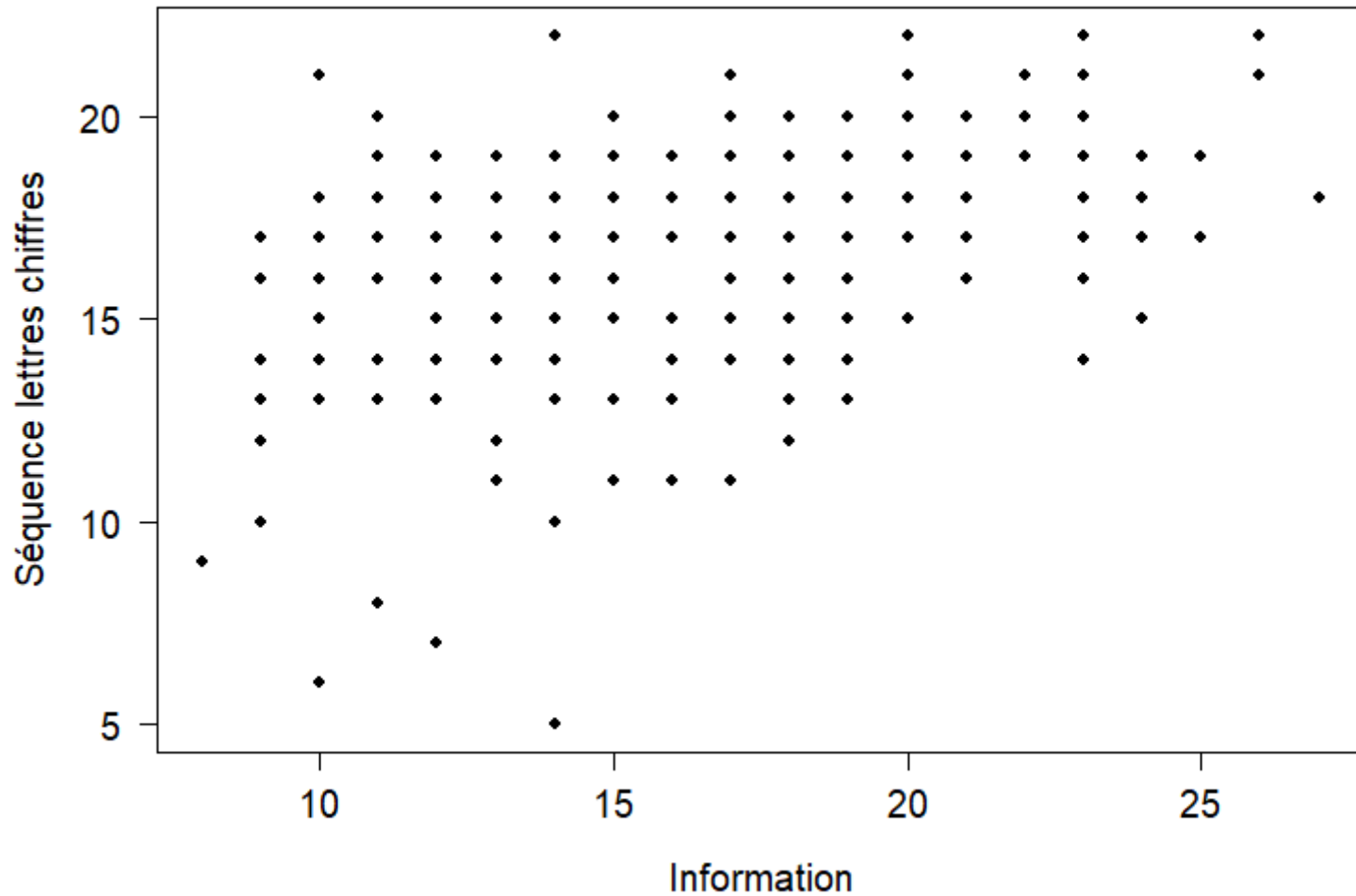
$N = 249$

Scores bruts uniquement : pas de scores normalisés !



	IDC	MAT	ARI	CUB	MDC	COD	SLC	BAR	VOC	SYM	CIM	RVE	SIM	INF
IDC	0.000	0.000	0.000	0.000	0	0.000	0.139	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MAT	0.128	0.000	0.070	0.087	0	0.000	0.259	0.036	0.000	0.089	0.132	0.050	0.000	0.000
ARI	0.294	0.271	0.000	0.213	0	0.066	0.517	0.204	0.110	0.334	0.255	0.215	0.079	0.000
CUB	0.126	0.200	0.115	0.000	0	0.000	0.322	0.144	0.000	0.222	0.248	0.058	0.000	0.000
MDC	0.573	0.505	0.514	0.448	0	0.277	0.784	0.445	0.285	0.563	0.514	0.406	0.248	0.008
COD	0.410	0.351	0.344	0.341	0	0.000	0.523	0.440	0.164	0.507	0.409	0.274	0.146	0.000
SLC	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BAR	0.074	0.030	0.006	0.041	0	0.000	0.132	0.000	0.000	0.176	0.074	0.000	0.000	0.000
VOC	0.448	0.410	0.395	0.351	0	0.192	0.615	0.390	0.000	0.453	0.456	0.409	0.273	0.084
SYM	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CIM	0.000	0.000	0.000	0.000	0	0.000	0.035	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RVE	0.374	0.326	0.291	0.243	0	0.067	0.496	0.254	0.212	0.332	0.382	0.000	0.188	0.002
SIM	0.462	0.417	0.386	0.376	0	0.191	0.593	0.366	0.288	0.452	0.474	0.402	0.000	0.069
INF	0.592	0.581	0.584	0.549	0	0.360	0.769	0.533	0.493	0.623	0.649	0.590	0.470	0.000

Exemples d'application

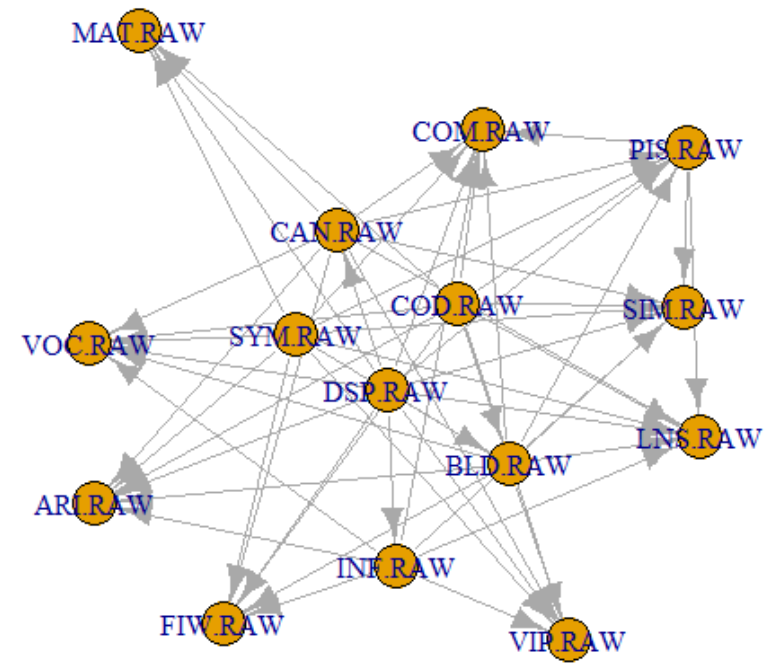


Exemples d'application

Données WISC5 (Pearson)

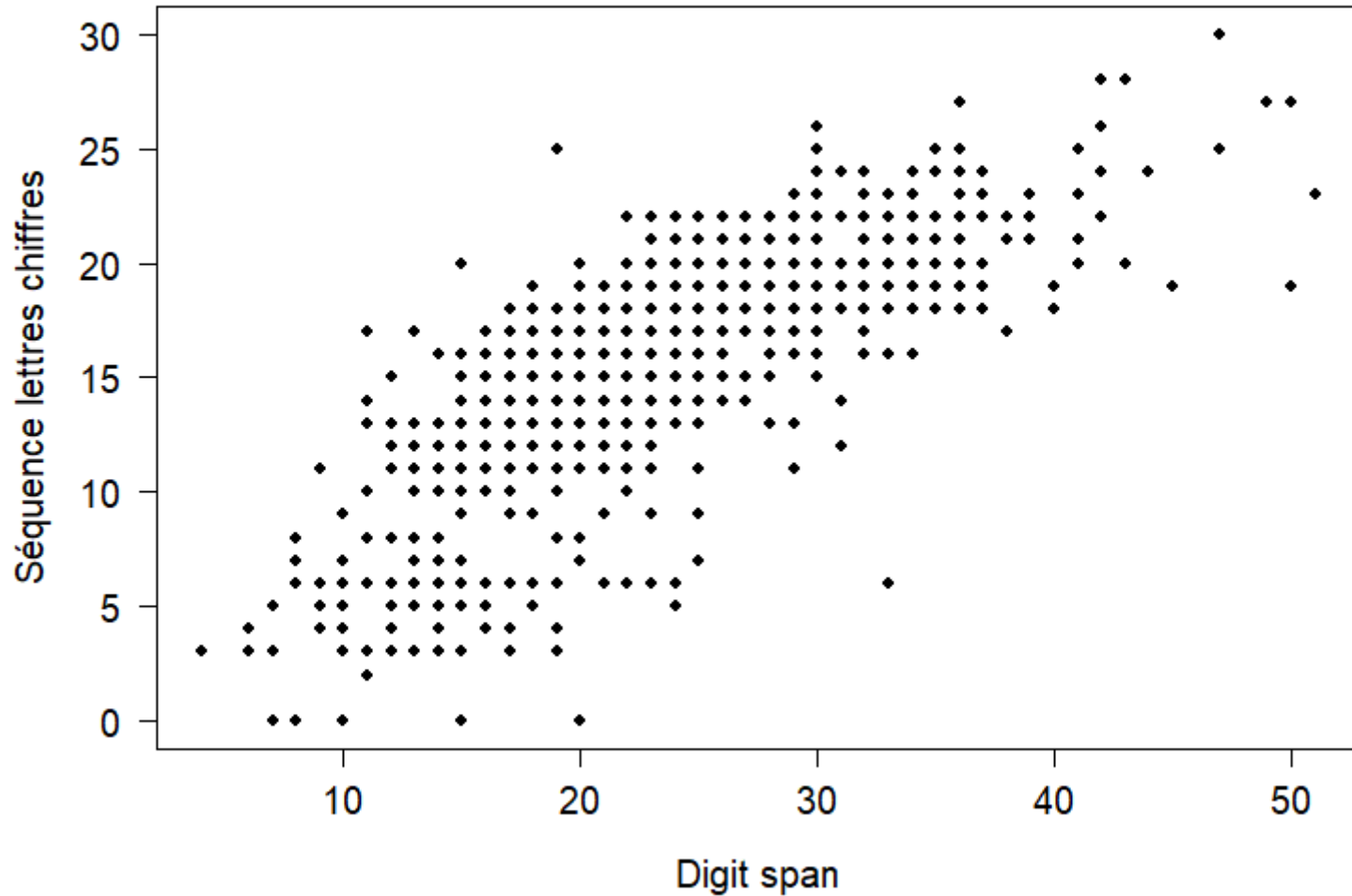
$N = 1049$

Scores bruts uniquement : pas de scores normalisés !



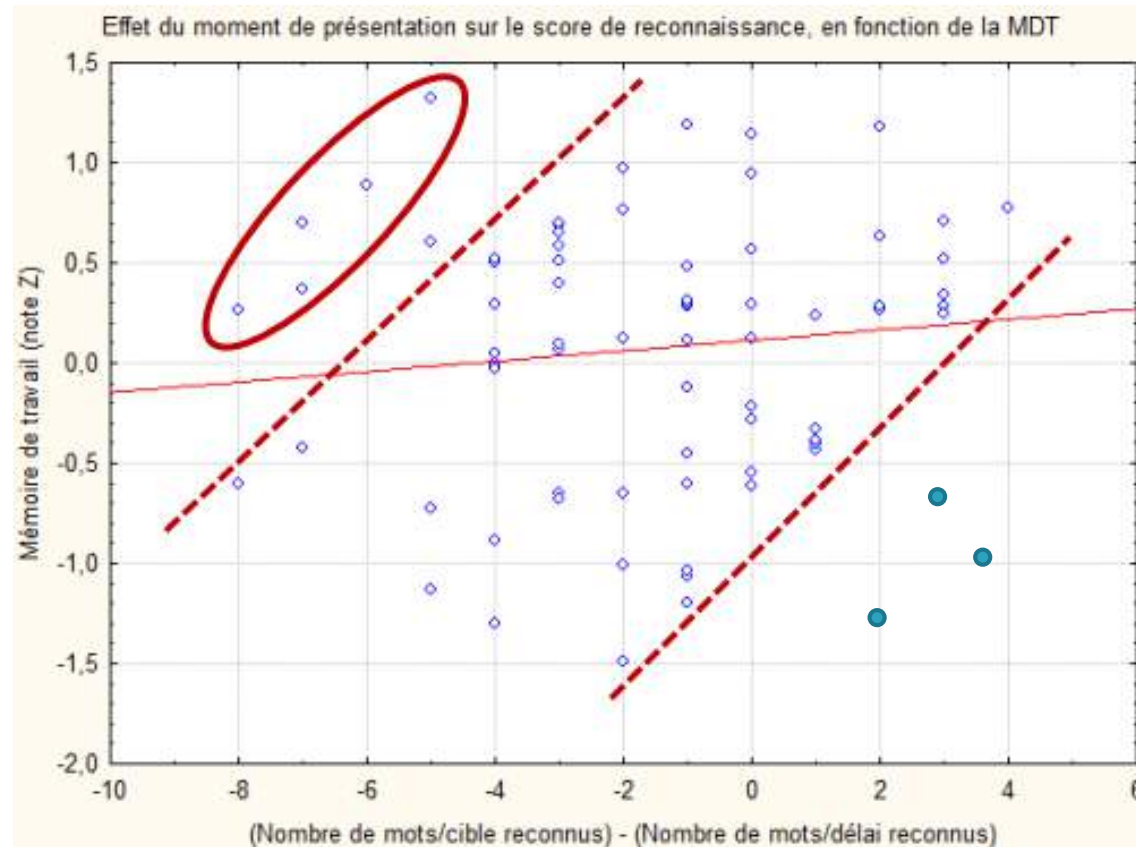
	BLD.RAW	SIM.RAW	MAT.RAW	DSP.RAW	COD.RAW	VOC.RAW	FIW.RAW	VIP.RAW	PIS.RAW	SYM.RAW	INF.RAW	LNS.RAW	CAN.RAW	COM.RAW	ARI.RAW
BLD.RAW	0.000	0.533	0.430	0	0.052	0.512	0.482	0.483	0.411	0.025	0.326	0.508	0.304	0.506	0.445
SIM.RAW	0.109	0.000	0.158	0	0.000	0.352	0.236	0.168	0.132	0.000	0.121	0.289	0.019	0.317	0.208
MAT.RAW	0.243	0.413	0.000	0	0.000	0.389	0.360	0.330	0.279	0.000	0.196	0.396	0.141	0.391	0.318
DSP.RAW	0.502	0.682	0.576	0	0.159	0.666	0.631	0.586	0.563	0.142	0.474	0.699	0.417	0.673	0.604
COD.RAW	0.431	0.595	0.474	0	0.000	0.582	0.533	0.500	0.488	0.173	0.383	0.592	0.418	0.605	0.510
VOC.RAW	0.151	0.401	0.202	0	0.000	0.000	0.287	0.211	0.171	0.000	0.170	0.321	0.046	0.376	0.244
FIW.RAW	0.189	0.366	0.249	0	0.000	0.356	0.000	0.284	0.225	0.000	0.144	0.338	0.073	0.333	0.280
VIP.RAW	0.282	0.400	0.306	0	0.000	0.375	0.369	0.000	0.291	0.000	0.190	0.386	0.150	0.379	0.315
PIS.RAW	0.274	0.439	0.331	0	0.000	0.413	0.388	0.365	0.000	0.000	0.223	0.444	0.216	0.422	0.348
SYM.RAW	0.376	0.537	0.429	0	0.136	0.517	0.481	0.461	0.435	0.000	0.313	0.544	0.358	0.550	0.453
INF.RAW	0.365	0.566	0.428	0	0.041	0.555	0.483	0.444	0.404	0.004	0.000	0.535	0.299	0.553	0.461
LNS.RAW	0.000	0.192	0.056	0	0.000	0.164	0.109	0.070	0.056	0.000	0.000	0.000	0.000	0.164	0.104
CAN.RAW	0.365	0.527	0.406	0	0.103	0.502	0.456	0.433	0.418	0.073	0.320	0.521	0.000	0.523	0.442
COM.RAW	0.019	0.274	0.087	0	0.000	0.284	0.142	0.098	0.067	0.000	0.042	0.202	0.000	0.000	0.131
ARI.RAW	0.241	0.425	0.298	0	0.000	0.403	0.370	0.320	0.276	0.000	0.216	0.420	0.167	0.409	0.000

Exemples d'application



Exemples d'application

Epilogue : l'implication 15 ans plus tard



$\phi = 56.22$

entropique = 26.07

implifiance = 42.11

$\psi = 61.94$

EII = 0.54

sEII = 0.10

$ps > .20$

➔ Vulnérabilité des liens implicatifs sur données numériques aux valeurs rares

