

# Pour la différence entre deux proportions jumelées, un nouveau test, plus valide et plus puissant

## A new standard normal-based test for the difference between paired proportions to supersede obsolete McNemar-like and other indirect procedures

Louis Laurencelle <sup>a</sup>

<sup>a</sup>Université du Québec à Trois-Rivières

**Abstract** ■ Contrarily to the 6-faced dice or the head/tail coin with their a priori fixed probability values, proportions used in applied research are generally based on heterogeneous and inconstant sources, the mathematical binomial model suiting them only as a first approximation. Moreover, the shape of their distributions is strongly tied to each proportion's mean value, a fact that rules out a direct binomial calculation for comparing them and assessing their difference. When the compared proportions are paired, i.e. based on the same sources, the awkwardness of the binomial solution simply jumps skyward, their proposed implementations being doubtful and their exegeses warped and indirect. Quinn McNemar's 1947 chi-squared solution, simple and straightforward, has long won users' adhesion, however it is based on the sole subset of option changing data pairs, putting aside all stable ones. We hereby describe a new, fully documented procedure for testing the difference between two paired proportions. It is anchored on the normal probability model and uses the Fisher-Zubin-Anscombe binomial-to-normal transformation. It is shown to be more precise and more powerful than the previous indirect and convoluted approaches, and it links empirical proportions to the full set of linear variables qualified for standard normal-based analyses, including ANOVA.

**Keywords** ■ McNemar, difference between paired proportions, alternative test.

[louis.laurencelle@gmail.com](mailto:louis.laurencelle@gmail.com)

[10.20982/tqmp.19.3.p254](https://doi.org/10.20982/tqmp.19.3.p254)

**Acting Editor** ■ Denis Cousineau (Université d'Ottawa)

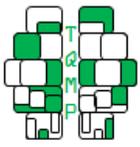
### Introduction

La variable binaire, l'ingrédient individuel entrant dans la formation d'une proportion, peut prendre l'un de deux états, par exemple « vrai / faux », « réussi / raté », « favorable / défavorable » etc., états numériquement symbolisés par les chiffres 1 et 0, le 1 dénotant le succès du résultat probable, le 0 l'échec. Considérée statistiquement, chaque source de données a une probabilité individuelle de  $\pi_i$  de produire le succès (1) et une de  $1 - \pi_i$  d'y échouer (0).

La proportion  $p$ , un agrégat de  $n$  données binaires  $y_i$ , en représente en fait la moyenne arithmétique, soit  $p = \sum_{i=1}^n y_i / n$ , où  $y_i$  ( $= 0$  ou  $1$ ) est la donnée de la source  $i$  basée

sur sa propension  $\pi_i$  vers un succès. Si la source concernée est de propension absolument stable, tel dans le cas d'une pièce de monnaie non biaisée qui a la probabilité de  $\pi = 1/2$  de produire le côté « face », ou celle d'un dé régulier à 6 faces qui produira le côté « 6 » avec probabilité  $\pi = 1/6$ , les différentes données  $y_i$  qui en résulteront répondront toutes à la force de propension constante  $\pi$  et la probabilité d'obtenir  $x$  valeurs 1, soit  $x = \sum_{i=1}^n y_i$ , obéira alors à la loi binomiale, de longue histoire (voir p. ex. STIGLER, 1986), nous y revenons plus loin.

En statistique appliquée, par contraste avec le domaine mathématique pur, l'application la plus courante de la proportion concerne des sources vivantes, chacune étant sus-



ceptible de changer, notamment changer de niveau de propension individuel  $\pi_i$  dans un contexte donné, et ces sources présentent de l'une à l'autre des propensions plus ou moins différentes. Pour ces proportions, la stricte interprétation de la loi binomiale classique, telle que déterminée par le paramètre-clé  $\pi$ , est utopique : cette loi peut néanmoins servir d'outil sous-tendant un modèle de calcul approximatif dont le résultat restera à vérifier empiriquement.

**Le changement de proportions.** Nonobstant les indications données ci-dessus, il est souvent intéressant de mesurer sous deux contextes différents les mêmes sources et déterminer si, d'un contexte à l'autre et selon les conditions présidant à chaque évaluation, les propensions ont changé, plus précisément, si l'hypothèse statistique de non changement,  $H_0 : \pi_1 = \pi_2$  tient ou non la route. Dans un tel cas, le test portera sur la différence  $p_1 - p_2$  des proportions établies en deux fois sur le même groupe de sources. Le cas plus courant de la différence entre deux proportions indépendantes, c.-à-d. entre deux groupes, a déjà été étudié (Laurencelle, 2005, 2017; Laurencelle et Cousineau, 2023) et sera brièvement mentionné plus tard.

Les cas concernés par l'étude du changement de proportions réfèrent au degré de stabilité d'option du répondant, représentée ici par le paramètre « constance » ou  $C$ , lequel ira de 0 (variation d'option au pur hasard) à  $C < 1$ , où la valeur « 1 » dénote une tendance absolue à conserver la même réponse  $y$  que celle donnée à la première évaluation. La constance 1 est ici exclue à l'échelle de l'échantillon observé, étant donné qu'elle assure d'emblée des séries comparées identiques, donc sans variance ni information utile à tirer par analyse.

**Contexte statistique.** Les proportions observées obéissent donc à peu près à la loi binomiale,<sup>1</sup> cependant leurs paramètres principaux  $\pi_1$  et  $\pi_2$  en sont inconnus. Posons un modèle stochastique possible basé sur la variable aléatoire uniforme  $u \sim U(0, 1)$  et la fonction de réponse  $f(u, \pi) = 1$  si  $u < \pi$  ou  $0$  sinon, nous aurions ainsi :

- a)  $\pi_1$  : propension individuelle initiale de  $y_1$  pour 1, soit  $y_1 = f(u, \pi_1)$ ;
- b)  $\pi_2$  : propension individuelle subséquente pour  $y_2$ , sous condition  $C$ ;
- c)  $C$  [constance d'option] : degré (de 0 à 1) de stabilité d'option pour  $y_2$ , sous : (a)  $y_2 \leftarrow y_1$  si  $u < C$ , sinon  $y_2 = f(u, \pi_2)$  [nouvelle propension], soit : (b) si  $C = 0$ , soit (b)  $y_2 = 1 - y_1$  [virement d'option], soit (c) hasard pur :  $y_2 = 1$  si  $u < 1/2$  ou 0 sinon). Ces trois modalités d'instabilité des options ont servi à couvrir une partie

du domaine interprétable des variations d'option imaginables.

Dans plusieurs contextes d'application, si  $y_{1i}$ , l'observation 1 de la source  $i$ , relève nécessairement d'un modèle binomial à paramètre  $\pi_{1i}$ , applicable à l'observation 1, ce paramètre est susceptible d'avoir évolué à l'observation 2 chez cette source, notamment après un changement du contexte de mesure ou suite à une influence temporelle ou expérimentale. Ce changement, toutefois, peut être quelque peu, voire plus que peu différent chez les  $n-1$  autres sources observées. Les proportions calculées seront donc généralement établies sur des mixtures paramétriques internes, mixtures par ailleurs généralement indéterminables.<sup>2</sup> La variable totalisée  $x = \sum_{i=1}^n y_i$  au numérateur de la proportion n'obéira plus à la loi binomiale stricte, et les aspérités de cette loi (mises en évidence notamment par sa variance  $\sigma^2$  et son fort indice d'asymétrie  $\gamma_1$ , voir plus loin) seront épointées, pavant la route vers une solution basée sur la loi normale, comme nous verrons bientôt.

L'importance contextuelle de ces considérations et leur possible obstruction à l'élaboration d'un jugement de probabilité mathématiquement rigoureux tiennent au fait que la variable proportion  $p = x/n$  montre autant de distributions différentes que le nombre de valeurs possibles des paramètres  $\pi_i$  et  $C$  qui lui sont sous-jacents.

### Les deux problèmes à considérer

Le défi ici est de comparer deux statistiques globales,  $p_1 = x_1/n$  et  $p_2 = x_2/n$ , chacune émanant idéalement d'une loi binomiale commune sous même taille échantillonnale  $n$  et amalgames paramétriques  $\pi_1$  et  $\pi_2$  inconnus, l'hypothèse nulle supposant l'égalité  $\pi_1 = \pi_2$ . Le but de la comparaison est de vérifier si la différence  $p_2 - p_1$  observée sur  $n$  sources permet d'affirmer qu'il existe une différence  $\pi_1 - \pi_2$  crédible chez la population d'où elles proviennent. Or, la loi binomiale qu'on aimerait vouloir invoquer ici dépend en tous points des paramètres  $n$  et  $\pi$  qui la définissent. Ces paramètres ont des effets majeurs non seulement sur le niveau moyen ( $\mu$ ) de la variable, mais aussi sur sa variance ( $\sigma^2$ ), qui réagit en V inversé selon  $\pi$ , son asymétrie ( $\gamma_1$ ) corrélée négativement avec  $\pi$ , et son aplatissement ( $\gamma_2$ ) qui élève des branches leptokurtiques aux deux extrémités du domaine de  $\pi$  : voir les formules (5) plus loin. Ces caractéristiques et leurs valeurs, qui dépendent lourdement du paramètre  $\pi$ , gouvernent essentiellement les calculs de probabilité. Or, au contraire du lancer du dé ou des jeux mécaniques d'un casino, contextes de probabilité où la valeur de  $\pi$  est définie, vérifiable et fixe, ce n'est pas le cas

1. La propension  $\pi_i$  de chaque source (vivante) étant instable en même temps que variable d'une source à l'autre, le modèle binomial (voir (4) plus bas) fait ici figure d'approximation, la distribution réelle du  $p$  résultante étant une mixture de binomiales instables.

2. Certains contextes d'observation scientifique peuvent exister dans lesquels la variabilité interne (entre les éléments de la proportion) de la valeur des paramètres  $\pi$  peut être réduite ou structurellement nulle (p. ex. dans les mesures physiques en usinage ou les tests de fonctionnement de produits industriels).

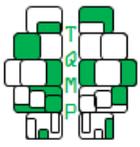


TABLEAU 1 ■ Dénombrément des paires de réponses binaires « 0 » ou « 1 » de  $n$  sources

		Avant		
		0	1	
Après	0	$a$ $n_{0,0}$	$b$ $n_{0,1}$	$n_0$
	1	$c$ $n_{1,0}$	$d$ $n_{1,1}$	$n_1$
		$n_{0\cdot}$	$n_{1\cdot}$	$n$

des données d'études auxquelles nous nous intéressons ici. Le second problème émane du fait que les sources de données - soit par exemple les personnes qui doivent choisir entre deux options, les animaux de laboratoire dont on cote le succès de performance, les patients dont on teste l'évolution, en fait tout ce qui peut être interrogé et réinterrogé dans la biosphère vis-à-vis des deux options résultantes, la disposition  $\pi_i$  de chacun pour l'une ou l'autre des options présentées est flottante et elle varie presque certainement d'une source à l'autre, chacune ayant sa propension personnelle, instable et changeante. L'ensemble des options prises par un groupe de sources est donc issu d'un mélange confus de propensions  $\pi_i$ , et il faut admettre que le calcul de probabilité en est, ici comme ailleurs, une approximation plus ou moins grossière.<sup>3</sup> Comment alors juger de la significativité du changement de réponse chez un groupe de sources observées deux fois, autrement dit, comment tester statistiquement la différence entre deux proportions jumelées?

### Fronts d'attaque du problème dans la littérature

#### La solution de McNemar, 1947 (voir aussi Siegel, 1956, ou Siegel et Castellan, 1988)

Posons d'emblée un tableau-modèle de données semblable à celui utilisé par McNemar (1947) et répandu dans toute la documentation traitant du sujet (voir Tableau 1). Les  $n$  sources produisant chacune deux données  $y_i$  se répartissent en quatre groupes de réponses, soit  $n_{0,0}$  (0 pour la première fois et 0 pour la deuxième), puis  $n_{0,1}$ ,  $n_{1,0}$  et  $n_{1,1}$ ; ces nombres sont aussi notés respectivement  $a$ ,  $b$ ,  $c$  et  $d$ . Les sommes marginales sont, par exemple,  $n_{0\cdot} = n_{0,0} + n_{0,1}$  (tous ceux qui ont donné un « 0 » en première réponse) et  $n_{1\cdot} = n_{1,0} + n_{1,1}$ . Sur les  $n$  sources ayant opté deux fois,  $n_{0,0} + n_{1,1}$  ( $= a + d$ ) ont chaque fois donné la même réponse, alors que  $n' = n_{0,1} + n_{1,0}$  ( $= b + c$ ) en ont changé. Finalement,  $p_1$  et  $p_2$  regroupent chacune les sources ayant produit l'option 1 les deux fois, plus celles l'ayant donné

une seule fois, respectivement pour  $p_1$  et  $p_2$ . soit :

$$p_1 = (n_{0,1} + n_{1,1})/n = (b + d)/n;$$

et

$$p_2 = (n_{1,0} + n_{1,1})/n = (c + d)/n. \tag{1}$$

Les proportions  $p_1$  et  $p_2$  sont jumelées (et non mutuellement indépendantes), la quantité  $n_{1,1}$  (ou  $d$ ) les associant numériquement.

McNemar (1947) a le premier<sup>4</sup> proposé une procédure pour asseoir la décision de changement ou de non changement concernant la différence  $p_2 - p_1$  obtenue en deux temps sur les mêmes  $n$  sources. Sa solution rappelle que la différence entre les proportions  $p_1$  et  $p_2$  est entièrement attribuable aux  $n'$  sources qui ont changé, soit les  $n_{0,1}$  qui sont passées de 0 à 1 et les  $n_{1,0}$  de 1 à 0. Sur la question de savoir si ces changements d'un côté ou de l'autre ont été produits au hasard, l'hypothèse nulle stipulera que chaque source a opté au hasard, selon une probabilité de  $\pi = 1/2$ , les effectifs tombant dans chaque option devant théoriquement approcher la valeur moyenne  $(n_{0,1} + n_{0,1})/2$ . L'écart par rapport à cette moyenne est-il significatif? Un calcul binomial simple permet d'y répondre ou, plus simplement encore, l'approximation normale

$$z = (|n_{0,1} - n_{1,0}| - 1) / \sqrt{n_{0,1} + n_{1,0}},$$

$$z \sim N(0, 1) \tag{2}$$

ou, identiquement :

$$z = (|b - c| - 1) / \sqrt{b + c},$$

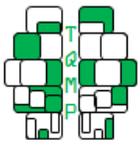
$$z \sim N(0, 1).$$

Cette formule d'estimation rejoint adéquatement une loi binomiale de paramètre  $\pi = 1/2$ , ce dès que  $n_{0,1} + n_{1,0}$  dépasse 8.

Le test de Wilcoxon (1945) sur les « différences de rangs avant-après signées », paru dans la foulée de la littérature pédagogique en statistique appliquée du XX<sup>e</sup> siècle, a donné lieu au dit « test des signes », basé sur les seuls

3. Cette « confusion » des valeurs de propriétés descriptives des sources est presque universelle dans le domaine des sources vivantes et concerne moult variables, comme le poids, la force, l'efficacité intellectuelle, etc., la plupart répondant assez bien au modèle de répartition gaussien (normal)... ce qui n'est pas le cas de la variable « proportion »!

4. Mais voir la référence à Sprent (1989), plus bas.



signes (+,-) des différences avant-après non nulles. Ce dernier test, faussement attribué à Wilcoxon, reproduit totalement le test binomial du test de McNemar. Sprent (1989) nous apprend aussi que le Londonien John Arbuthnot, médecin et mathématicien écossais bien connu, l'utilisait déjà en 1710 dans le cadre d'études d'ordre épidémiologique!

La « solution McNemar », précisément identique au test des signes, est présentée sous diverses formules à peu près équivalentes (Fagerland, Lydersen et Laake 2014); la suggestion de Liddell (1983) est du même acabit.

**Suissa et Shuster (1991) et d'autres**

Sous le titre « Tests recommandés et intervalles de confiance pour des proportions binomiales jumelées », Fagerland et cie (2014), outre les variantes du test de McNemar signalées ci-dessus (parmi lesquelles ils auraient pu inclure celle de Lu 2010), présentent un modèle de solution mieux articulé, celui de Suissa et Shuster (1991). Ce modèle s'appuie sur un calcul multinomial dans lequel la différence «c-b», basée sur les b+c changements d'option, est enchâssée dans la combinatoire du tableau 2 x 2 global, exploitant donc toutes ses n données. Nonobstant son mérite, le test de Suissa et Shuster ne fait qu'approcher la solution cherchée, la différence globale entre les proportions p1 et p2 échappant encore au test : ce qui est testé est le caractère exceptionnel de la différence «c-b» dans son contexte multinomial global. De plus, le test de la différence «c-b» se fait en explorant son domaine de probabilité multinomiale, notamment en variant les valeurs possibles des cellules a et d, ce qui a pour effet ordinaire d'émousser l'acuité du test et de diminuer sa puissance, comme c'est le cas des estimations de type bayésien (voir aussi Laurencelle 2017).

La documentation présente d'autres approches situées dans la périphérie du problème étudié, par exemple celle de Liang et Zeger (1988) qui développent un modèle log-linéaire complexe pour le traitement de proportions corrélées, ou celles que citent Johnson, Kotz et Kemp (1972, p. 148-150), notamment une chaîne Markov fournissant un calcul de probabilité des variables binomiales corrélées.

Pouvons-nous mettre sur pied un test direct de la différence entre deux proportions corrélées, soit p1 - p2 = [(b+d) - (c+d)]/n, une procédure de test légitime, précise et puissante, d'application facile et qui s'annexe harmonieusement au répertoire des tests d'hypothèses statistiques à référence normale? C'est ce que nous proposons ici.

**Notre solution : le test normal de la différence entre deux moyennes corrélées**

**L'approche classique**

Le test classique qui compare deux moyennes répondant à la loi normale et est basé sur n paires de données X jumelées a pour formes possibles les deux principales expressions suivantes, soit :

$$t = \sqrt{n}(\bar{X}_1 - \bar{X}_2) / \sqrt{(s_1^2 + s_2^2)(1 - r_{1,2})/n}, \quad (3a)$$
$$dl = n - 1,$$

une forme générale qui admet une légère inégalité des variances estimées, la variance d'erreur résultante étant réduite grâce à la corrélation positive r1,2 entre les deux séries Xi, épongeant ainsi une bonne part de l'effet d'hétérogénéité des sources, ou la forme :

$$t = \sqrt{n} \times \bar{d} / s_d, \quad (3b)$$
$$dl = n - 1,$$

celle-ci à calcul expéditif (voir p. ex. Sokal et Rohlf, 1995, p. 354), basée sur les différences individuelles di = x1i - x2i avec leur moyenne d̄ et leur écart-type sd, le contrôle de l'hétérogénéité se faisant cette fois via la covariance implique s1,2 = r1,2 · s1 · s2.

**Admissibilité de la variable « p » pour l'approche normale classique**

Peut-on appliquer l'approche normale classique au test de proportions? Formellement, la proportion p = x/n (0 ≤ x ≤ n) répond à la loi binomiale B(n, π), soit :

$$pr(x|n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad (4)$$

les moments de p (Johnson, Kotz et Kemp, 1992) étant

$$\mu' = \pi;$$
$$\sigma^2 = n\pi(1 - \pi);$$
$$\gamma_1 = (1 - 2\pi)/\sigma;$$
$$\gamma_2 = (1 - 6\pi(1 - \pi))/\sigma^2. \quad (5)$$

On constate d'abord que, si l'espérance (μ') de p est π, sa juste cible, sa variance σ² bouge fortement avec π, allant d'un minimum près de 0 pour π → 0 et π → 1 pour un maximum de n/4 pour π = 1/2. Le moment 3 (γ1), indice d'asymétrie de p, fait de même : symétrique (γ1 = 0) lorsque π = 1/2, p devient contre-asymétrique selon π, l'aile droite

5. « Leptokurtique », des mots grecs leptο (λεπτο : mince, maigre) et kurtose (κνρτωση : courbure) signifie pointue, élançé (vers le haut), tandis que son opposé, platykurtique, de platy (πλατυν : large, étendu) signifie bien étalé, jusqu'à plat.

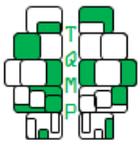


TABLEAU 2 ■ Valeur minimale de  $\pi$  pour obtenir  $G(p') = |\gamma_1| + \frac{1}{2} |\gamma_2| < 0,5$  selon  $n$

$n$	$p$ brute	$p'$ (Anscombe)
5	0,31	0,22
10	0,20	0,12
25	0,13	0,06
50	0,09	0,03
100	0,05	0,02
500	0,02	0,01
1000	0,01	0,01

Note. Extrait de Laurencelle (2021a, p. 169)

de sa distribution s'allongeant vers la droite ( $\gamma_1 > 0$ ) pour  $\pi \rightarrow 0$ , et vice-versa ( $\gamma_1 < 0$ ) pour  $\pi \rightarrow 1$ . L'indice d'aplatissement  $\gamma_2$  lui-même est affecté : légèrement platykurtique pour  $\pi$  près de  $1/2$ , la distribution de  $p$  devient leptokurtique pour  $\pi$  approchant l'une ou l'autre de ses valeurs extrêmes.<sup>5</sup>

Les caractéristiques distributionnelles de la simple proportion  $p$  disqualifient généralement l'application de procédures à base normale. Même lorsque la taille échantillonnaire  $n$  augmente, atténuant les déformations de symétrie et d'étalement de la distribution de  $p$ , la variance elle-même, de calcul  $n \cdot \pi \cdot (1 - \pi)$ , demeure totalement affectée par la variation de  $\pi$ , un effet qui mettrait le test de différence entre deux (ou plusieurs) estimateurs  $p$  en contradiction avec la condition d'homoscédasticité de l'approche normale classique.

**La solution Fisher (1922) / Zubin (1935) / Anscombe (1948)**

Et si la variable  $p$  pouvait être numériquement rapprochée de la loi normale? C'est à R. A. Fisher (1890-1962) que nous devons la plus grande part des concepts, méthodologies et techniques statistiques utilisés aujourd'hui pour la recherche expérimentale. Fisher valorisait particulièrement les approches statistiques assises sur la loi normale en raison de sa grande souplesse et de son statut asymptotique, plusieurs autres lois de probabilité convergeant vers elle avec l'accroissement de la taille échantillonnaire. Par exemple, après avoir découvert la distribution exacte du  $r$  sous  $\rho = 0$ , puis celle convenant à tout  $\rho$ , Fisher a proposé l'excellente transformation  $r' = \tanh^{-1}(r)$  qui normalise presque parfaitement la distribution de  $r$ , tous moments considérés. Sur cet exemple et à la suggestion de Fisher, Zubin en 1945 propose pour la proportion  $p$  la fonction trigonométrique  $p' = \sin^{-1} \sqrt{p}$ , et il montre que sa variance, peu affectée par  $\pi$ , est égale à  $1/\sqrt{4n}$  en première approximation. Anscombe, en 1948, suggère à son tour la formule améliorée

$$\sin^{-1} \sqrt{\left(x + \frac{3}{8}\right) / \left(n + \frac{3}{4}\right)},$$

cette fois de variance  $1/\sqrt{4n+2}$ . La fonction d'Anscombe et d'autres fonctions similaires sont minutieusement examinées dans Laurencelle (2021a), lequel rapporte que « La transformation angulaire rapproche indéniablement la variable  $p$  d'une répartition normale, de façon quasi complète à  $n = 100$  pour tout  $\pi$  et dès  $n = 25$  pour  $0,02 \leq \pi \leq 0,95$ . » (ibid., p. 169). Le tableau 2, emprunté à Laurencelle (ibid.), fait état du bon comportement « normal » de  $p'$  selon la taille échantillonnaire utilisée : on y trouve le  $n$  requis pour que l'indice équilibré  $G(x)$  (asymétrie + aplatissement) de non-normalité n'excède pas 0,5.

L'autre atout déterminant en faveur de  $p'$  est la relative et asymptotique insensibilité à  $\pi$  de sa variance. L'étude citée et ses compléments (Laurencelle, 2021b, Laurencelle et Cousineau, 2023) plus axés vers les applications documentent la validité et l'efficacité pratique de cette approche.

En écho au modèle normal classique (3a) pour la comparaison de deux moyennes jumelées (et corrélées), nous proposons la formule générale (notation en référence au Tableau 1) :

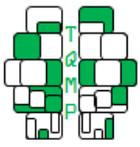
$$z [f(p_1) - f(p_2)] = \frac{f(x_1, n) - f(x_2, n)}{\sqrt{2\sigma_n^2(1 - \phi_{1,2})}}, \quad (6)$$

où  $x_1 = c + d$ ;  $x_2 = b + d$ ;  $n = a + b + c + d$ ;  $\sigma_n^2 = 1/4n$ ; et

$$f(x, n) = \sin^{-1} \sqrt{\frac{x + 3/8}{n + 3/4}};$$
$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Rappelons que  $\phi$  (phi) désigne le coefficient de corrélation applicable aux données binaires 0/1; ce coefficient est mathématiquement identique au  $r$  de Pearson pour ce contexte.

**Note :** Le lecteur aura remarqué que la variance de  $1/4n$  retenue pour la fonction d'Anscombe diffère de celle,  $1/(4n+2)$ , justifiée par l'auteur. L'une des raisons avancées par ce dernier se rapporte à la comparaison de proportions dont



l'une serait basée sur un petit  $n$  (p. ex.  $n < 10$ ) et l'autre sur un plus grand, l'impact dépendant de cet écart étant ainsi corrigé. Or, dans le cas du test de proportions jumelées, les tailles  $n$  des deux proportions sont nécessairement égales et nos données de simulations nous ont montré que la réduction de variance amenée par la formule d'Anscombe provoque dans le test un sursaut de significativité pour des tailles  $n$  inférieures à 50, sursaut qui déforme nettement la croissance par ailleurs monotone de la significativité du test (cf. Figures 2 plus loin).

### Quelques exemples comparatifs

Les bilans statistiques plus bas illustrent à la fois les calculs et les propriétés comparatives des procédures de test appliquées.

**Tableau 3a.** Les  $n = 21$  données paires du Tableau 3 (en page suivante) sont tirées de Fagerland et cie (2014, p. 2851, Table II); elles peuvent aussi être représentées sous forme de paires associées brutes, comme suit :

Avant : 1 1 0 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1  
Après : 1 1 0 1 1 0 1 1 0 0 0 1 0 1 1 1 0 1 0 1 1

Le lecteur peut vérifier que les moyennes Avant ( $p_1$ ) et Après ( $p_2$ ) sont 0,905 et 0,619, valeurs qu'on retrouve par  $p_1 = (c + d)/n = 19/21$  et  $p_2 = (b + d)/n = 13/21$ . La différence  $p_1 - p_2$  entre ces moyennes est simplement  $(c - b)/n = 6/21 \approx 0,286$ . Nous référant aux formules (6), nous obtenons les valeurs transformées :

$$p'_1 = \sin^{-1} \sqrt{(19 + 3/8) / (21 + 3/4)} \approx 1,234$$
$$p'_2 = \sin^{-1} \sqrt{(13 + 3/8) / (21 + 3/4)} \approx 0,901,$$

et

$$\sigma_n^2 = 1/(4 \times 21) \approx 0,0119,$$
$$\phi = (1 \times 12 - 1 \times 7) / \sqrt{2 \times 19 \times 8 \times 13} \approx 0,0795,$$

d'où le calcul du test  $z$ (Anscombe) plus bas.

### Bilan statistique (3a)

$p_1$ vs $p_2$	0,9048 vs 0,6190
$z$ (Anscombe cf. (6))	2,2470
pr(ANS)	0,0246 (probabilité bilatérale)
$\phi$ ou $r$	0,0795
$z$ (McNemar cf. (1))	1,7678
pr(McNemar cf. (1))	0,0771
pr(McNemar cf. Binomiale)	0,0703 (calcul binomial exact)
pr(Suissa et Shuster)	0,0353

### Bilan statistique (3b)

$p_1$ vs $p_2$	0,2381 vs 0,2857
$z$ (Anscombe, cf. (6))	-2,0382
pr(ANS)	0,0415 (probabilité bilatérale)
$\phi$ ou $r$	0,8839
$z$ (McNemar cf. (1))	-1,5000
pr(McNemar cf. (1))	0,1336
pr(McNemar cf. Binomiale)	0,1250 (calcul binomial exact)

### Bilan statistique (3c)

$p_1$ vs $p_2$	0,4854 vs 0,5146
$z$ (Anscombe, cf. (6))	-2,4762
pr(ANS)	0,0133 (probabilité bilatérale)
$\phi$ ou $r$	0,9434
$z$ (McNemar cf. (1))	-2,0412
pr(McNemar cf. (1))	0,0412
pr(McNemar cf. Binomiale)	0,0313 (calcul binomial exact)

**Au tableau 3a,** les 21 fréquences inscrites sont réparties entre paires stables ( $a + d = 13$ ) et paires changeantes ( $b + c = 8$ ), la portion relativement modeste de données stables expliquant la mince corrélation positive  $\phi = 0,0795$  obtenue. Le test de McNemar sur la comparaison  $c (= 7)$  vs  $b (= 1)$  n'atteint pas la significativité bilatérale au seuil de 0,05, laquelle est atteinte (0,0246) par le test  $z$  classique sur deux moyennes jumelées accommodé à la transformation Anscombe. À noter, la probabilité de 0,0353 obtenue pour ce tableau de Fagerland et cie (2014) via l'approche multinomiale de Suissa et Shuster (1991), une approche, rappelons-le, qui cible encore spécifiquement la différence  $c - b$ , cette fois en tenant compte du contexte global des  $n$  données.

**Au tableau 3b,** le grand déséquilibre des quelques changements ( $n_{1,0} = 0$  vs  $n_{0,1} = 4$ ) ne parvient pas à toucher la significativité, qu'obtient encore notre  $z$  classique, largement grâce à l'épuration de variance d'erreur par la forte corrélation présente ( $\phi = 0,8839$ ). Une fois la corrélation ignorée ou mise à zéro – comme si nous avions affaire à deux groupes distincts de 42 sources chacun –, le  $z$  tomberait à 0,6945 pour une probabilité normale de 0,4874.

**Au tableau 3c,** Le déséquilibre 0 vs 6 des 6 changements permet au test de McNemar d'atteindre tout juste la significativité (0,0412 ou 0,0313), que le test classique obtient haut-la-main (0,0133), encore une fois grâce à la corrélation présente : sous corrélation  $\phi = 0$ , nous aurions eu  $z = 0,5891$  pour  $pr \approx 0,5558$ .

### Justifications paramétriques de la solution proposée

Les sections suivantes rapportent succinctement le résultat d'examen comparatifs formels des statistiques utilisées dans la présente étude, soit la proportion brute  $p$  et deux variantes de la fonction d'Anscombe, la plupart des examens se rapportant à la validité statistique du test (6) proposé.

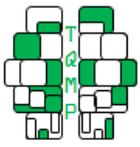
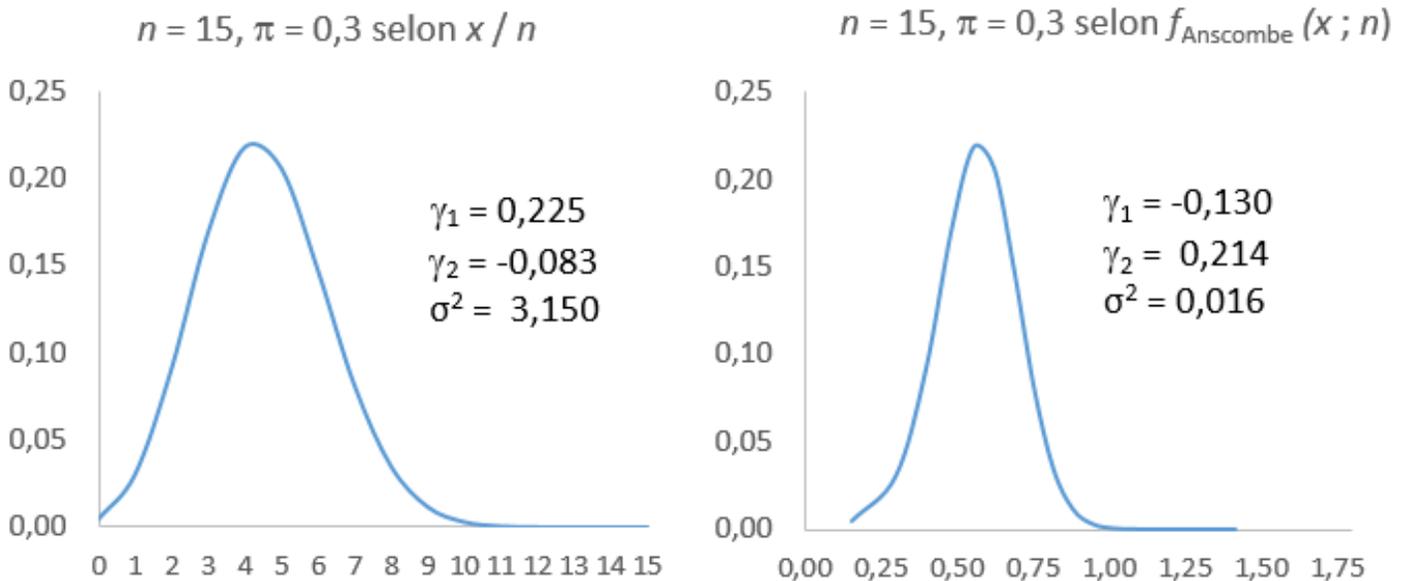


TABLEAU 3 ■ Trois exemples de matrices sommatives de données binaires pairées.

<p style="text-align: center;"><i>Tableau 3a</i></p> <table border="0" style="margin-left: auto; margin-right: auto;"> <tr><td colspan="2"></td><td colspan="2" style="text-align: center;">Avant</td><td></td></tr> <tr><td colspan="2"></td><td style="text-align: center;">0</td><td style="text-align: center;">1</td><td></td></tr> <tr><td rowspan="4" style="vertical-align: middle;">Après</td><td style="text-align: center;">0</td><td style="border: 1px solid green; padding: 2px;"><i>a</i></td><td style="border: 1px solid green; padding: 2px;"><i>b</i></td><td style="text-align: right;">2</td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;">1</td><td style="border: 1px solid green; padding: 2px;">1</td><td></td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;"><i>c</i></td><td style="border: 1px solid green; padding: 2px;"><i>d</i></td><td style="text-align: right;">19</td></tr> <tr><td></td><td style="border: 1px solid green; padding: 2px;">7</td><td style="border: 1px solid green; padding: 2px;">12</td><td></td></tr> <tr><td></td><td style="text-align: center;">8</td><td style="text-align: center;">13</td><td style="text-align: right;">21</td></tr> </table>			Avant					0	1		Après	0	<i>a</i>	<i>b</i>	2	1	1	1		1	<i>c</i>	<i>d</i>	19		7	12			8	13	21	<p style="text-align: center;"><i>Tableau 3b</i></p> <table border="0" style="margin-left: auto; margin-right: auto;"> <tr><td colspan="2"></td><td colspan="2" style="text-align: center;">Avant</td><td></td></tr> <tr><td colspan="2"></td><td style="text-align: center;">0</td><td style="text-align: center;">1</td><td></td></tr> <tr><td rowspan="4" style="vertical-align: middle;">Après</td><td style="text-align: center;">0</td><td style="border: 1px solid green; padding: 2px;"><i>a</i></td><td style="border: 1px solid green; padding: 2px;"><i>b</i></td><td style="text-align: right;">64</td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;">60</td><td style="border: 1px solid green; padding: 2px;">4</td><td></td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;"><i>c</i></td><td style="border: 1px solid green; padding: 2px;"><i>d</i></td><td style="text-align: right;">20</td></tr> <tr><td></td><td style="border: 1px solid green; padding: 2px;">0</td><td style="border: 1px solid green; padding: 2px;">20</td><td></td></tr> <tr><td></td><td style="text-align: center;">60</td><td style="text-align: center;">24</td><td style="text-align: right;">84</td></tr> </table>			Avant					0	1		Après	0	<i>a</i>	<i>b</i>	64	1	60	4		1	<i>c</i>	<i>d</i>	20		0	20			60	24	84	<p style="text-align: center;"><i>Tableau 3c</i></p> <table border="0" style="margin-left: auto; margin-right: auto;"> <tr><td colspan="2"></td><td colspan="2" style="text-align: center;">Avant</td><td></td></tr> <tr><td colspan="2"></td><td style="text-align: center;">0</td><td style="text-align: center;">1</td><td></td></tr> <tr><td rowspan="4" style="vertical-align: middle;">Après</td><td style="text-align: center;">0</td><td style="border: 1px solid green; padding: 2px;"><i>a</i></td><td style="border: 1px solid green; padding: 2px;"><i>b</i></td><td style="text-align: right;">106</td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;">100</td><td style="border: 1px solid green; padding: 2px;">6</td><td></td></tr> <tr><td style="text-align: center;">1</td><td style="border: 1px solid green; padding: 2px;"><i>c</i></td><td style="border: 1px solid green; padding: 2px;"><i>d</i></td><td style="text-align: right;">100</td></tr> <tr><td></td><td style="border: 1px solid green; padding: 2px;">0</td><td style="border: 1px solid green; padding: 2px;">100</td><td></td></tr> <tr><td></td><td style="text-align: center;">100</td><td style="text-align: center;">106</td><td style="text-align: right;">206</td></tr> </table>			Avant					0	1		Après	0	<i>a</i>	<i>b</i>	106	1	100	6		1	<i>c</i>	<i>d</i>	100		0	100			100	106	206
		Avant																																																																																													
		0	1																																																																																												
Après	0	<i>a</i>	<i>b</i>	2																																																																																											
	1	1	1																																																																																												
	1	<i>c</i>	<i>d</i>	19																																																																																											
		7	12																																																																																												
	8	13	21																																																																																												
		Avant																																																																																													
		0	1																																																																																												
Après	0	<i>a</i>	<i>b</i>	64																																																																																											
	1	60	4																																																																																												
	1	<i>c</i>	<i>d</i>	20																																																																																											
		0	20																																																																																												
	60	24	84																																																																																												
		Avant																																																																																													
		0	1																																																																																												
Après	0	<i>a</i>	<i>b</i>	106																																																																																											
	1	100	6																																																																																												
	1	<i>c</i>	<i>d</i>	100																																																																																											
		0	100																																																																																												
	100	106	206																																																																																												

FIGURE 1 ■ Profils de densité des proportions  $p = x/n$  brute (gauche) vs  $p'$  transformée (droite).



**La normalité.** La figure 1 visualise le profil distributionnel d'une proportion selon son expression brute  $x/n$ , à gauche et, à droite, selon la transformation d'Anscombe (1948)  $\sin^{-1} \sqrt{(x+3/8)/(n+3/4)}$ ; l'illustration porte ici sur un échantillon de taille modeste ( $n = 15$ ). Les valeurs paramétriques données à la figure de gauche obéissent aux formules (5), celles de la figure de droite étant directement obtenues par un calcul exact des moments.

Rappelons que les indices de forme de la distribution, c.-à-d. l'asymétrie  $\gamma_1$  et la kurtose  $\gamma_2$  (dite aussi voûture ou aplatissement) de la loi normale sont tous deux fixés à 0. Ici, pour la variable binomiale  $p$ , le paramètre  $\pi = 0,3$ , parce qu'inférieur à sa valeur médiane 0,5, provoque une asymétrie positive, sa kurtose associée étant presque normale; sa variance  $\sigma^2$ , qui avec  $n = 15$  peut varier de 0 à 3,75 selon  $\pi$  par sa formule (5), donne ici  $\sigma^2 = 3,15$ . Quant au profil distributionnel de la variable transformée, l'asymétrie est

comme toujours à la fois inversée et réduite en valeur par rapport à celle de la binomiale  $p$  correspondante : cette réduction est un effet typique de la transformation. Une légère leptokurtose apparaît dans le cas présent. Rappelons que ces indices de forme décroissent pour les deux variables en fonction de la taille  $n$ , le tandem  $\gamma_1$  et  $\gamma_2$  se rapprochant progressivement de 0, leur valeur 'normale'. On remarquera aussi la variance applicable à la série transformée selon Anscombe, soit 0,01619, déjà presque égale à sa valeur asymptotique théorique, soit  $1/(4 \times 15 + 2) \approx 0,01613$ .

**La précision du test.** Sous l'hypothèse nulle  $H_0 : \pi_1 - \pi_2 = 0$ , le test respecte-t-il son seuil prescrit de significativité  $\alpha$ ? Pour vérifier cette condition de validité du test, nous avons réalisé différentes études de type Monte Carlo. Les résultats les plus parlants sont illustrés dans les figures 2 : elles font voir le chemin de probabilité que parcourt chacun des tests comparés pour rejoindre le palier de signifi-

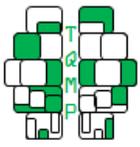
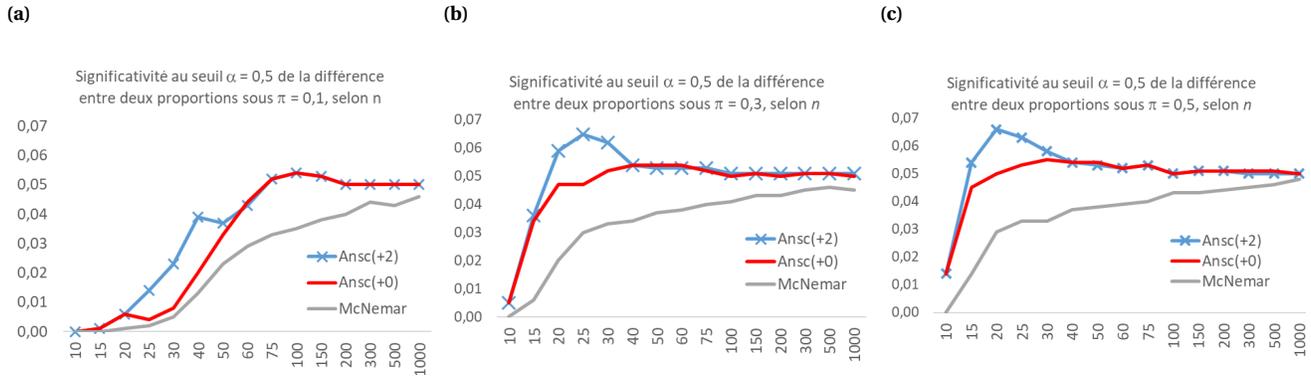


FIGURE 2 ■ Profils de significativité de trois tests de différence entre proportions jumelées selon la taille  $n$  pour  $\pi = 0, 1, 0,3$  et  $0,5$



cativité prescrit : le seuil de significativité alpha ( $\alpha$ ) prescrit est de 0,05, évalué en mode bilatéral. Les tests comparés sont notre solution (6) notée ici Ansc(+0) indiquant une variance fixée à  $1/(4n)$ , sa variante Ansc(+2) utilisant la variance  $\sigma^2 = 1/(4n + 2)$  que propose Anscombe (1948), puis le test de McNemar (2). Les données graphiques sont basées sur  $10^6$  échantillons Monte Carlo.

Les figures 2a à 2c font voir chacune la montée des probabilités des trois tests vers le niveau  $\alpha$  en fonction de la taille d'échantillon. Les élévations apparaissent de plus en plus abruptes dans l'ordre  $\pi = 0, 1 < 0,3 < 0,5$ , c.-à-d. à mesure que  $\pi$  approche de  $1/2$  et la variable approche de la distribution normale. On notera aussi les écarts extravagants de la courbe notée Ansc(+2), celle utilisant la variance proposée par Anscombe (1948), le léger renchérissement du test (pour des tailles  $n$  modestes) faisant basculer la probabilité calculée au-dessus de sa valeur réelle. Les données de précision associées à des valeurs  $\pi$  de 0,9 et 0,7 reproduisent respectivement celles de nos  $\pi$  de 0,1 et 0,3. Enfin, le rendement du test de McNemar ressort manifestement lacunaire.

En résumé, le seuil  $\alpha = 0,05$  de notre test (6) applicable à la comparaison de deux proportions jumelées n'est garanti qu'à partir d'un échantillon de sources assez nourries. Par exemple, pour le cas  $\pi_1 = 0,1$  vs  $\pi_2 = 0,1$ , le test n'entre dans l'environnement du seuil qu'à partir de  $n \approx 70$  et il s'y case à partir de  $n \approx 200$ ; ces nombres régressent quelque peu pour  $\pi = 0,3$  et  $0,5$ , la stabilisation advenant vers  $n = 100$ .

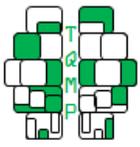
**La puissance du test.**

La puissance statistique d'un test est une fonction de deux paramètres, la différence entre les valeurs paramétriques comparées et la taille échantillonnale  $n$  : la significativité du ou des tests en lice devra déborder le seuil d'exceptionnalité  $\alpha$ , puis croître vers la certitude 1 selon les

paramètres mentionnés, cela à un régime qui dépend du test utilisé. Nous produisons ici sous forme graphique une sélection de résultats que nous jugeons illustratifs, ceux-ci encore basés sur  $10^6$  échantillons Monte Carlo. Sous la condition  $C = 0$ , la source, ayant produit sa première option sous la propension de grandeur  $\pi_1$  (c.-à-d. qu'elle optera d'abord pour  $y_1 = 1$  dans une proportion de cas proche de  $\pi_1$ , ou pour  $y_1 = 0$  sinon), va produire indépendamment sa seconde option  $y_2$  sous  $\pi_2$ , cette situation ne générant pas de vraie corrélation entre les séries d'options 1 et 2 produites. Sous le  $C = 0,8$  appliqué dans notre deuxième condition, la source répétera avec probabilité 0,8 au temps 2 la valeur  $y$  donnée au temps 1, créant ainsi corrélation (positive) entre les séries qui deviennent alors effectivement jumelées, sinon elle optera pour un nouveau  $y_2$  avec probabilité  $\pi_2$ .

Les huit figures présentées autorisent à formuler trois constats sur les mérites relatifs des tests comparés, le premier étant la forte concordance entre les performances des trois tests, le second est la légère mais persistante infériorité du test de McNemar. Quant à la comparaison entre le test appliquant la version originale de la fonction Anscombe, avec variance fixée à  $1/(4n + 2)$ , et celle de  $1/4n$  (donnée par Zubin) que nous proposons, le gonflement artificiel de la « variance Anscombe » pointe le nez dans les quatre graphiques de gauche, une légère surenchère de significativité apparaissant pour les tailles  $n$  inférieures à 50.

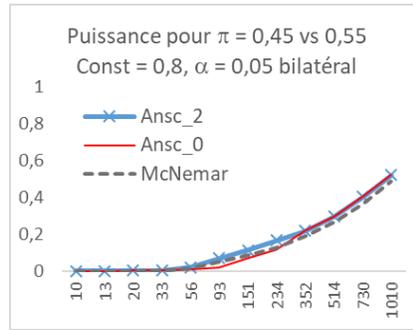
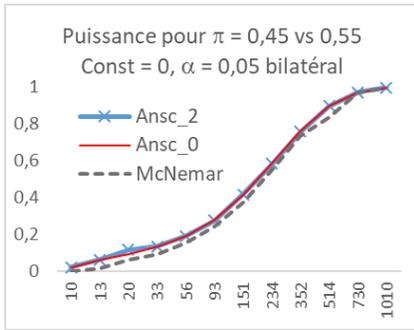
Le constat le plus remarquable ici est l'affaissement de puissance dans les graphiques avec 'constance'  $C = 0,8$ . Cette « perte de puissance » apparente liée à une condition de constance positive découle du fait que, pour leur seconde option  $y_2$ , les sources ont une probabilité  $C$  définie de conserver pour  $y_2$  leur réponse  $y_1$  relative à  $\pi_1$  pour le deuxième temps, à défaut de quoi, sous probabilité  $1 - C$ , leur réponse  $y_2$  sera régie par la propension  $\pi_2$ . Prenons l'exemple des deux figures 3g et 3h, mettant en jeu  $\pi_1 = 0,4$



**FIGURE 3** ■ Profils de puissance de trois tests de différence entre proportions jumelées en fonction de la taille  $n$  pour 4 couples  $\pi_1$  vs  $\pi_2$  et 2 valeurs de constance  $C$ .

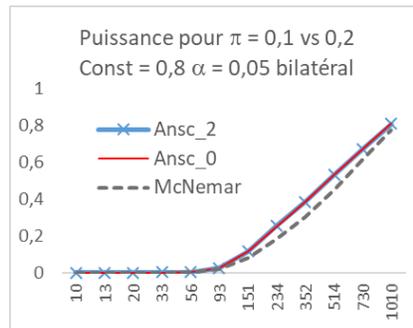
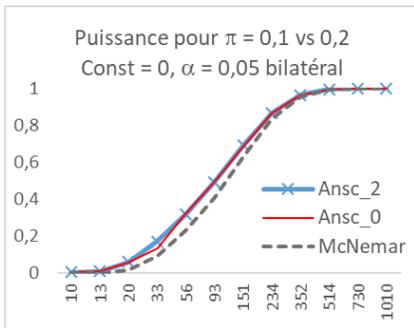
(a) – constance = 0

(b) – constance = 0,8



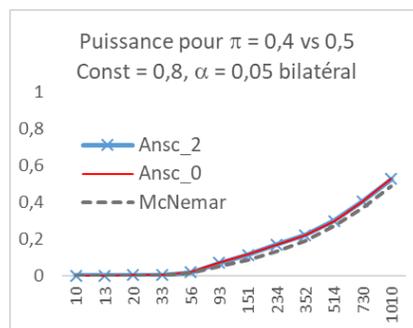
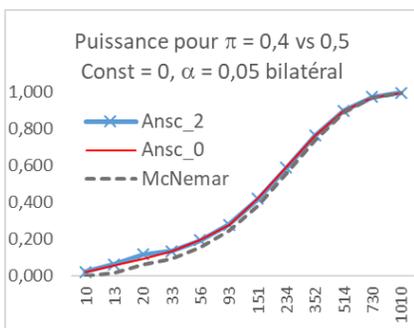
(c) – constance = 0

(d) – constance = 0,8



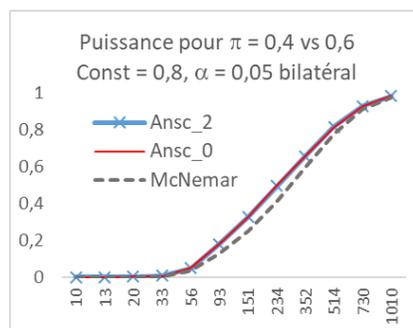
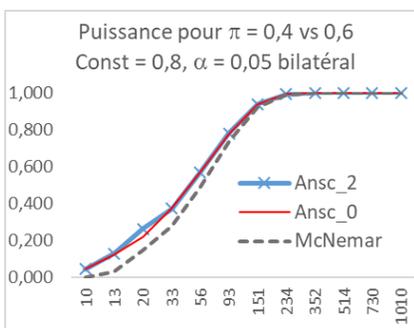
(e) – constance = 0

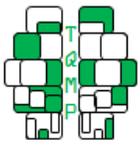
(f) – constance = 0,8



(g) – constance = 0

(h) – constance = 0,8





et  $\pi_2 = 0,6$  : la constance de 0,8 appliquée pour le graphique de droite exprime une « fidélité » d'option ou d'opinion vraisemblablement plus proche de celle qu'on rencontre dans les contextes d'études courants, les sondages d'opinion par exemple. Calculés selon notre formule (6) sur  $10^6$  échantillons Monte Carlo de  $n = 100$  sources chacun, nous obtenons pour la colonne de gauche :

$$\begin{aligned} C &= 0,0, \\ p_1 &= 0,400(\pm 0,049), \\ p_2 &= 0,600(\pm 0,049), \\ \phi &= 0,000(0,100), \\ Pu &= 0,810, \end{aligned}$$

et pour la colonne de droite :

$$\begin{aligned} C &= 0,8, \\ p_1 &= 0,401(\pm 0,049), \\ p_2 &= 0,439(\pm 0,050), \\ \phi &= 0,787(0,060), \\ Pu &= 0,203. \end{aligned}$$

La puissance effective pour les cas de seconde votation fidèle reflète, comme on le voit, l'effet d'un rapprochement des proportions (ou moyennes) comparées. Dans les cas d'indépendance entre les premiers et seconds votes ( $C = 0$ ), nous constatons l'observance de la différence des conditions de chaque contexte de probabilité, soit  $p_2 - p_1 \approx \pi_1 - \pi_2$ , tributaire d'une corrélation  $\phi$  à peu près nulle entre les sources étudiées. Sous une constance d'option d'ordre 0,8, la corrélation ( $\phi = 0,787$ ) joue : en même temps, cette corrélation réduit l'impact du second contexte (régé par  $\pi_2$ ) sur les résultats, soit  $p_2 \approx \pi_1 + (\pi_2 - \pi_1)(1 - C)$ , ou  $p_2 - p_1 \approx (\pi_2 - \pi_1)(1 - C)$ , la différence attendue étant amortie par la fidélité des sources. Ici, le  $p_2$ , paramétré proche du 0,6 ( $= \pi_2$ ), frôle plutôt la valeur  $0,4 + (0,6 - 0,4) \times (1 - 0,8) = 0,440$ .<sup>6</sup> Finalement, il est intéressant de noter que, dans le cas du présent exemple admettant une significativité de 0,203, l'omission de la corrélation ( $1 - \phi$ ) dans la formule (6) écraserait la significativité de 0,203 du test à la valeur 0,002.

La condition de  $C = 0$  cas correspond pratiquement, le lecteur l'autre compris, à la situation comparant deux proportions indépendantes, c.-à-d. des proportions établies sur deux groupes de sources de tailles  $n$  quelconques. Les références à caractère général déjà fournies couvrent ces situations, mais le lecteur sera peut-être intéressé à deux tests dits « exacts » de la différence  $p_1 - p_2$ , ceux-là à calcul proprement binomial, le premier (Liddell, 1978) s'appuyant sur le principe du maximum de vraisemblance, le

second (Laurencelle, 2017) étant de type bayésien. Une comparaison de la performance de notre test inspiré de la transformation d'Anscombe avec celles de ces deux compétiteurs, mathématiquement plus rigoureux mais plus lourds, devrait valoir l'effort.

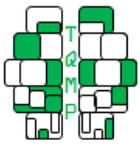
## Épilogue

Jusqu'à récemment, le test de McNemar (1947) a été la solution dominante, voire la seule solution répandue pour déterminer si la différence entre deux proportions basées sur le même échantillon de sources est statistiquement sérieuse, une solution initialement popularisée par Siegel (1956), qui vérifie si, pour les seules sources ayant changé d'option, l'une des deux options choisies domine statistiquement l'autre. La prise en compte de toutes les données, incluant celles des sources répétant deux fois la même option, n'avait pas trouvé de solution souple, simple et généralisable. La solution proposée ici, justifiée par la transformation normalisante inspirée de Fisher (1922) et Zubin (1935) et basée sur Anscombe (1948), intègre le test des proportions dans l'univers des tests à base normale et, plus généralement, l'inclut dans le giron de l'analyse de variance (Laurencelle, 2021b, Laurencelle et Cousineau, 2023), avec toutes ses ramifications techniques. L'approche normalisante pour le test sur deux proportions, jumelées ou non, est certes approximative, tout comme l'approche binomiale le serait aussi, ainsi que nous l'avons argumenté. Ce caractère approximatif est dans la nature même de la loi normale associée au théorème central-limite, ce qui, loin d'amoinrir sa valeur instrumentale, en justifie plutôt le mérite. Enfin, le test normalisé, basé sur la formule (6), affiche des propriétés de précision et de puissance qui le placent en tête de ses braves compétiteurs.

## Références

- ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35, 246-254.
- FAGERLAND, M. W., LYDERSEN, S., & LAAKE, P. (2014). Recommended tests and confidence intervals for paired binomial proportions. *Statistics in medicine*, 33, 2850-2875.
- FISHER, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society (Edinburgh)*, 42, 321-341.
- JOHNSON, N. L., KOTZ, S., & KEMP, A. W. (1992). *Univariate discrete distributions (2e édition)*. Wiley.
- LAURENCELLE, L. (2005). Le test binomial exact de la différence entre deux proportions et ses approximations. *Lettres Statistiques*, 12, 1-14.

6. La belle correspondance déductive des données attendues avec les paramètres ( $\pi_2$  et  $\phi$ ) du modèle de simulation n'est pas facilement inversible en raison de la relation réciproque entre la différence  $\pi_2 - \pi_1$  et  $\phi$ . Posant  $\pi_1 \approx p_1$  et  $C \approx \phi$  nous avons  $\pi_2 \approx p_1 + (\hat{\pi}_2 - p_1)(1 - \phi)$ , puis  $\hat{\pi}_2 \approx p_1 + (p_2 - p_1)/(1 - \phi) \approx 0,583$  (vs  $\pi_2 = 0,600$ ). L'estimation sous  $n < \infty$  est naturellement imprécise, De plus, cette correspondance approximative ne nous informe en rien sur la significativité, c.-à-d. l'exceptionnalité de la différence obtenue.



- LAURENCELLE, L. (2017). Le test binomial exact de la différence entre deux proportions et ses approximations [The exact binomial test between two proportions and its approximations]. *The Quantitative Methods for Psychology*, 13, 141-147.
- LAURENCELLE, L. (2021a). Étude comparative de trois fonctions de transformation normalisante d'une proportion (Anscombe, 1948 – Freeman-Tukey, 1950 – Chanter, 1975). *The Quantitative Methods for Psychology*, 17, 166-197.
- LAURENCELLE, L. (2021b). Le traitement statistique des proportions incluant l'analyse de variance, avec des exemples / The statistical handling of proportions including analysis of variance, with worked out examples. *The Quantitative Methods for Psychology*, 17, 272-285.
- LAURENCELLE, L., & COUSINEAU, D. (2023). Analysis of proportions using arcsine transform with any experimental design. *Frontiers (Front Psychol., 13)*, 1045436. doi : [10.3389/fpsyg.2022.1045436](https://doi.org/10.3389/fpsyg.2022.1045436).
- LIANG, K. L., & ZEGER, S. L. (1988). On the use of concordant pairs in matched case-control studies. *Biometrics*, 44, 1145-1156.
- LIANG, K.-Y., & ZEGER, S. L. (1988). On the use of concordant pairs in matched case-control studies. *Biometrics*, 44, 1145-1156.
- LIDDELL, D. (1978). Practical tests of 2x2 contingency tables. *The Statistician*, 25, 295-304.
- LIDDELL, F. D. K. (1983). Simplified analysis of case-referent studies : matched pairs; dichotomous exposure. *Journal of Epidemiology and Community Health*, 37, 82-84.
- LU, Y. (2010). A revised version of McNemar's test for paired binary data. *Communications in Statistics – Theory and Methods*, 39, 3525-3539. doi : [10.1080/03610920903289218](https://doi.org/10.1080/03610920903289218).
- MCNEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 13-158.
- SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.
- SIEGEL, S., & CASTELLAN, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.
- SOKAL, R. R., & ROHLF, F. J. (1995). *Biometry (3rd edition)*. Freeman.
- SPRENT, P. (1989). *Applied Nonparametric Statistical Methods (2e édition)*. Chapman & Hall.
- STIGLER, S. M. (1986). *The history of statistics : The measurement of uncertainty before 1900*. Harvard University Press.
- SUISSA, S., & SHUSTER, J. J. (1991). The 2 x 2 matched-pairs trial : Exact unconditional design and analysis. *Biometrics*, 47, 361-372.
- WILCOXON, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1. doi : [10.2307/3001968](https://doi.org/10.2307/3001968).
- ZUBIN, J. (1935). Note on a transformation function for proportions and percentages. *Journal of Applied Psychology*, 19, 2123-220.

## Citation

- LAURENCELLE, L. (2023). Pour la différence entre deux proportions jumelées, un nouveau test, plus valide et plus puissant [A new standard normal-based test for the difference between paired proportions to supersede obsolete McNemar-like and other indirect procedures]. *The Quantitative Methods for Psychology*, 19(3), 254-264. doi : [10.20982/tqmp.19.3.p254](https://doi.org/10.20982/tqmp.19.3.p254).

Copyright © 2023, LAURENCELLE. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 14/07/2023 ~ Accepted: 04/08/2023