



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

Louis Laurencelle ^a

^aUniversité du Québec à Trois-Rivières

Abstract ■ The simple proportion, $p = x/n$, a pervasive statistical tool equally used in public and academic research areas, is in the literature still short of the necessary analytical implements needed for full-scale statistical treatments. This is partly ascribable to its discrete numerical character, but it proceeds mostly from its other distributional properties: its variance is tied up with its expectation, and its density shows a strong U-shaped asymmetry reactive to its π parameter, the habitual normal-based analytical procedures thus being contra-indicated. We here revive the Fisher-Yates angular transformation of the proportion, $y(x, n) = \sin^{-1}\sqrt{x/n}$, heralded for its π -independent variance and smoothed-out non-normality, and put to trial three of its improved descendants (Anscombe 1948, Tukey & Freeman 1950, Chanter 1975). Following a would-be thorough study of the three y functions retained (bias, variance, precision, test accuracy, power), largely documented in Laurencelle's (2021a) investigation, we develop and illustrate the z test of significance on one proportion, the z test on the difference of two independent proportions, the analysis of variance of $k \geq 2$ independent proportions and finally the test of two and anova of k correlated proportions. A critical appraisal of the McNemar test for the difference of two correlated proportions is also essayed.

Keywords ■ Proportions, Transformation angulaire, Proportions corrélées, Analyse de variance, Test de McNemar ■ Proportions, Angular transformation, Correlated proportions, Analysis of variance, McNemar's test.

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

louis.laurencelle@gmail.com

[10.20982/tqmp.17.3.p272](https://doi.org/10.20982/tqmp.17.3.p272)

Introduction

Les tests sur les proportions, une statistique bien répandue à la fois dans le monde scientifique et dans le monde civil, sont pourtant restés les parents pauvres des applications statistiques, et ils restent mal servis dans les publications savantes, les outils disponibles étant rares et ne couvrant pas tous les besoins. Nous décrivons ici une approche, basée sur une « transformation angulaire » attribuée à Fisher et Yates, transformation qui permet en principe d'effectuer valablement le test de la différence entre deux proportions indépendantes de même que des analyses de variance (anova) simples et à plan factoriel.

Ce qui existe déjà. Soit une proportion $p = x/n$, calculée à partir de n éléments binaires $x_i = 0$ ou 1 et $x = \sum x_i$. Le test de l'hypothèse nulle $H_0 : \epsilon(p) = \pi$ (où ϵ dénote l'espérance) a sa solution exacte par un calcul de l'intégrale binomiale en plus de bénéficier de diverses approximations (LAURENCELLE & DUPUIS, 2002b, 2002a; HOWELL, YZERBYT & BESTGEN, 2008). Quant à la différence entre deux proportions, $p_1 - p_2 = x_1/n_1 - x_2/n_2$, et l'hypothèse nulle $H_0 : \epsilon(p_1) = \epsilon(p_2)$, on lui a proposé des approximations dites normales, la plus intéressante étant



le test :

$$z_{p_1-p_2} = \frac{x_1/n_1 - x_2/n_2}{\sqrt{\frac{X(N-X)}{N^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad (1)$$

où $X = x_1 + x_2$, $N = n_1 + n_2$, avec ou sans correction de continuité au numérateur; le test classique du khi-deux, sur tableau 2×2 , lui est identique, à la mise au carré près. Le lecteur s'intéressera peut-être à deux tests exacts de la différence $p_1 - p_2$, c.-à-d. des tests à calcul binomial exact, l'un de LIDDELL (1978), basé sur l'estimation par maximum de vraisemblance du paramètre π de référence, l'autre de LAURENCELLE (2005, 2017b, 2021a), de type bayésien, évalué sur le domaine paramétrique entier de π .

D'autres statistiques, basées celles-là sur des fonctions $f(x, n)$ différentes du couple (x, n) , ont aussi vu le jour et elles ont gagné du terrain dans les applications, notamment en méta-analyse et en biologie. Citons la fonction $\text{logit}(p) = \ln[p/(1 - p)]$ ou $\ln[x/(n - x)]$ et la fonction $\text{probit}(p) = \Phi^{-1}(p)$ (SOKAL & ROHLE, 1995), où $\Phi(z)$ est l'intégrale normale standard jusqu'à z , de même que différents indices élaborés pour la mesure en épidémiologie (voir p. ex. ANCELLE, 2017). Il y a enfin la dite « transformation angulaire » attribuée à R. A. Fisher et F. Yates (BARTLETT, 1947), celle applicable à une proportion, $\sin^{-1}\sqrt{p}$, soit l'angle au carré dont le sinus donne p . À l'instar de la transformation $\tanh^{-1}r$ de Fisher qui stabi-

lise près de $1/(n - 3)$ la variance échantillonnale du coefficient de corrélation tout en normalisant sa distribution, la transformation angulaire de p ramène sa variance près de $1/(4n)$ selon π . Rappelons que cette propriété de variance stable d'une statistique joue un rôle clé à la fois dans la combinaison de deux ou plusieurs estimations d'une proportion commune et, bien sûr, dans la validité d'application de tests tels que le t ou le z pour la différence entre deux proportions de même que l'analyse de variance, l'objet principal de notre propos.

Les 3 fonctions candidates. Le lecteur l'aura compris, nous ne considérerons ici que les statistiques de proportions qui satisfont minimalement la condition de stabilité (ou égalité approximative) de variance, telle que requise généralement pour l'analyse de variance, le test t de différence entre deux moyennes et dans maintes autres applications comme la méta-analyse et l'épidémiologie quantitative. En vertu de cette exigence, nous excluons bien sûr la proportion p brute, la transformation $\text{probit}(p)$ dont la variance imite celle de p , la transformation $\text{logit}(p)$ à variance de valeur à peu près réciproque de celle de p , ainsi que d'autres indices à étalement non contrôlé. Il nous reste donc la transformation angulaire, que quelques chercheurs ont tenté d'affiner afin notamment d'en stabiliser la variance sur l'étendue la plus complète de π et de corriger si possible le biais subsymptotique que son espérance présente pour de faibles valeurs de n . Les trois fonctions retenues sont :

Anscombe (1948)	$y_{ANS} = \sin^{-1} \sqrt{\frac{x + 3/8}{n + 3/4}};$	$\text{var}(y) \approx 1/(4n + 2)$	(2a)
-----------------	---	------------------------------------	------

Freeman et Tukey (1950)	$y_{F-T} = \sin^{-1} \sqrt{\frac{x}{n + 1}} + \sin^{-1} \sqrt{\frac{x + 1}{n + 1}};$	$\text{var}(x) \approx 1/(n + 1/2)$	(2b)
-------------------------	--	-------------------------------------	------

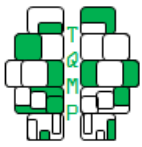
Chanter (1975)	$y_{CHA} = \sin^{-1} \sqrt{\frac{1}{4n} + \left(1 - \frac{1}{2n}\right) \frac{x}{n}};$	$\text{var}(x) \approx 1/4n,$	(2c)
----------------	--	-------------------------------	------

les angles étant exprimés en radians. Pour des conversions en degrés d'angle, les variances proposées doivent être multipliées par $(180/\pi)2 \approx 3283$.

La fonction $F-T$ (2b) est un calcul doublé d'à peu près la même expression, fournissant donc comme résultat

une valeur à peu près double des formules (2a) et (2c). À des fins de comparaison des formules, nous utiliserons la forme (2b'), notée $F-T/2$, obtenue en divisant la formule (2b) par 2, soit :

Freeman et Tukey'	$y_{F-T/2} = \left[\sin^{-1} \sqrt{\frac{x}{n + 1}} \right] + \left[\sin^{-1} \sqrt{\frac{x + 1}{n + 1}} \right] / 2;$	$\text{var}(x) \approx 1/(4n + 2)$	(2b')
-------------------	--	------------------------------------	-------



Les mérites communs de ces transformations. En plus de stabiliser la variance échantillonnale de $f(x, n)$ près de $1/(4n)$, la transformation angulaire, sous toutes ses formes, admet les valeurs limites $x = 0$ et $x = n$, ce qui n'est le cas ni du logit ni du probit. En outre et peut-être surtout, cette transformation trigonométrique décomprime progressivement les queues de la distribution, y restituant un début de symétrie et une enveloppe à peu près normalisée. La transformation a toutefois un inconvénient, celui d'un biais selon π , à savoir que, sous de faibles tailles n , elle tend à surestimer sa valeur stable (asymptotique) pour $\pi < 1/2$ et à la sous-estimer pour $\pi > 1/2$. Chanter a proposé sa formule (2c) expressément pour pallier ce problème. Enfin, le lecteur, après examen des formules (2a), (2b') et (2c), peut constater que les trois se confondent lorsque $n \rightarrow \infty$ et deviennent alors identiques à la formule de base, $\sin^{-1} \sqrt{x/n} = \sin^{-1} \sqrt{\pi}$, la valeur-cible asymptotique.

Notre « Étude comparative » (LAURENCELLE, 2021a) présente les résultats d'une étude des fonctions de transformation (2a), (2b') et (2c) quant au biais des fonctions sous petits n , la stabilisation de leur variance, l'approche de la forme normale, le respect du seuil de significativité (α) du test z sur une proportion et sur deux proportions indépendantes et, enfin, la puissance du test sur deux proportions. L'étude met aussi en compétition d'autres fonctions de calcul, telles que les tests normaux classiques sur une proportion et deux proportions (équivalents aux tests khi-deux correspondants) et les calculs binomiaux « exacts » de LAURENCELLE (2005, 2017b, 2021b) (voir aussi LIDDELL, 1978).

La suite de cet article se veut une illustration de l'usage de la transformation angulaire pour les tests z sur une proportion et sur deux proportions indépendantes ou corrélées ainsi que pour l'analyse de variance des proportions à devis simple sur groupes indépendants ou jumelés ou à devis factoriel. Pour ce contexte, nous recourons à tour de rôle aux trois fonctions étudiées, lesquelles sont généralement interchangeables mais possèdent des mérites parfois subtilement différents. Le lecteur, après examen de l'« Étude comparative » mentionnée, pourra se faire sa propre idée de chaque fonction, les trois venant à s'équivaloir aussitôt que les tailles n des échantillons concernés dans une analyse atteignent 100, voire 50. Nous concluons par l'énoncé de notre appréciation nuancée des vertus et des mérites de cette approche d'analyse.

Le test z sur une proportion

Exemple 1 (Prédire le comportement d'un dé). Robert, qui dit être « psychique », prétend pouvoir annoncer mieux que le hasard le comportement d'un dé à 6 faces. Pour le tester, nous lançons le dé en l'air 12 fois, à l'abri du regard

de notre devin. À chaque lancer, durant l'envolée du dé, Robert émet un chiffre, puis le chiffre apparaissant sur la face supérieure du dé retombé est noté, le tout à l'insu de Robert. Sa prédiction tombe juste 5 fois sur 12 : est-ce un résultat remarquable?

Le calcul exact. En stipulant que le dé est « honnête », chaque face ayant 1 chance sur 6 d'apparaître, la probabilité de deviner correctement 5 résultats ou plus est, par calcul binomial, de $P_{\text{Exact}} = Pr\{x > 5 | n = 10, \pi = 1/6\} = 0,03635 \approx 0,036$. Un résultat intéressant pour Robert mais non encore convaincant pour un esprit scientifique.

Les fonctions estimatives. Les estimations effectuées par le recours au modèle normal doivent compenser pour le caractère discontinu de la variable binomiale, ce en appliquant la correction de $\pm 1/2$ au numérateur de la proportion. C'est le cas des tests t et z classiques comme ce l'est de ceux recourant à la transformation « $\sin^{-1} \sqrt{p}$ » appliquée ici. Nos tests étant tous de mode unilatéral droit, nous appliquons partout la correction « $-1/2$ » au numérateur x de la proportion x/n .

Quant aux fonctions angulaires notées « y » sous étude, leur test emprunte la forme classique :

$$z(y) = \frac{y(x - 1/2, n) - y(\pi)}{\sqrt{\text{var}(y)}} \tag{3}$$

Pour illustration, nous recourons à la fonction y_{Cha} de Chanter (formule 2c), qui donne ici la variable :

$$y_{\text{Cha}} = \sin^{-1} \sqrt{\frac{1}{4 \times 12} + \left(1 - \frac{1}{2 \times 12}\right) \cdot \frac{5 - 1/2}{1} \cdot 2} \approx 0,6644,$$

puis le test :

$$z(y_{\text{Cha}}) = \frac{y_{\text{Cha}} - \sin^{-1} \sqrt{1/6}}{\sqrt{1/(4 \times 12)}} \approx 1,690$$

en utilisant $y(\pi = 1/6) \approx 0,4205$ et la variance appropriée, la probabilité normale étant $P_{\text{Cha}} = 1 - \Phi(z_{\text{Cha}}) \approx 0,046$. Le lecteur pourra aussi vérifier que

$$y_{\text{Ans}} \approx \sin^{-1} \sqrt{(5 - 1/2 + 3/8) / \left(12 + \frac{3}{4}\right)} \approx 0,6666,$$

d'où $z(y_{\text{Ans}}) \approx 1,740$, $P_{\text{Ans}} \approx 0,041$ et

$$y_{F-T/2} \approx \left[\sin^{-1} \sqrt{\frac{5 - 1/2}{12 + 1}} + \sin^{-1} \sqrt{\frac{5 - 1/2 + 1}{12 + 1}} \right] / 2 \approx 0,6686,$$

d'où $z(y_{F-T/2}) \approx 1,754$, $P_{F-T/2} \approx 0,040$.



Le test z classique,

$$z_{\text{Brute}} = \frac{x - n \cdot \pi \pm 1/2}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}, \tag{4}$$

convient mal au présent cas, la condition de validité prescrite, soit $n \times \min(\pi, 1 - \pi) \geq 10$, n'étant pas satisfaite. Nous calculons tout de même $z_{\text{Brute}} \approx 1,936$, pour une probabilité normale extrême de $P_{\text{Brute}} \approx 0,026$, relativement éloignée de P_{Exact} !

Les calculs exploitant les fonctions angulaires donnent ici tous la même sanction : significative au seuil $\alpha = 0,05$, non au seuil $\alpha = 0,01$, concordant avec la probabilité exacte de 0,036, ce par contraste avec le test z_{Brute} classique, occasionnellement désavantagé.¹ On constate donc un accord entre les fonctions, accord nuancé par de petites différences, un état de fait que notre étude extensive nous a montré fréquent, les différences étant vraisemblablement imputables au caractère discret, saltatoire, de la variable binomiale et au traitement spécifique que chaque fonction lui accorde.

Le test z de comparaison de deux proportions indépendantes

Exemple (L'efficacité du vaccin antigrippal chez les personnes âgées). Deux groupes de personnes âgées de 65 à 75 ans ont été recrutés dans trois résidences pour personnes autonomes, l'échantillonnage s'étant effectué au quasi-hasard. Les personnes souffrant de difficultés pulmonaires ou cardiaques permanentes ou transitoires n'étaient pas retenues. Informé(e) des buts de la recherche, chaque candidat ou candidate acceptait ou non d'y participer : on exhortait les uns à recevoir le vaccin antigrippal annuel, les autres à s'en abstenir. Sur 60 personnes des deux sexes, réparties entre les conditions avec versus sans vaccin, 26 sur 30 et 23 sur 30 ont signé leur formulaire de consentement. Le vaccin étant appliqué à la mi-automne, les 49 participants ont été suivis jusqu'au printemps prochain, soit à la fin mars. Dans le groupe « Avec vaccin », des symptômes grippaux sont apparus chez 5 des 26 personnes ($p_1 \approx 0,192$), alors que le groupe « Sans » en a enregistré 9 sur 23 ($p_2 \approx 0,391$). Le vaccin a-t-il un effet significatif ?

Le calcul exact. Pour fixer la probabilité extrême associée à cette différence, c.-à-d. calculer $Pr\{p_2 - p_1 \geq 0 | p_1 \sim B1(x_2, n_2, \pi), p_2 \sim B1(x_2, n_2, \pi)\}$, nous ne connaissons que deux solutions, celle de LIDDELL (1978), qui totalise les probabilités conjointes associées à tous les

couples (x_1, x_2) tels que $x_2/n_2 - x_1/n_1 \geq p_2 - p_1$, en invoquant la valeur paramétrique correspondant au maximum de vraisemblance conjointe des deux séries, soit $\hat{\pi} = (n_1 p_1 + n_2 p_2) / (n_1 + n_2) = (x_1 + x_2) / (n_1 + n_2)$. LAURENCELLE (2005, 2017b, 2021b) propose quant à lui d'examiner tout le domaine de valeurs de π (de 0 à 1), en pondérant chacune par sa probabilité conjoncturelle associée aux deux couples (x_i, n_i) observés. LAURENCELLE (2017b) montre que les deux solutions sont généralement très proches l'une de l'autre. Par le calcul de LAURENCELLE (2005, 2017b, 2021b), toujours en mode unilatéral, on obtient $P_{\text{Exact}} \approx 0,059$.

Les fonctions estimatives. Noter d'entrée de jeu que, dans le cas présent d'une différence entre deux proportions comme dans le cas à venir de l'analyse de variance, la différence, c.-à-d. la soustraction implicite des proportions l'une par rapport à l'autre, rend caduque la correction de continuité de $-1/2$ applicable au test d'une seule proportion :² nous ne l'avons donc pas appliquée.

La formule générale a la forme :

$$z(y_1 - y_2) = \frac{y_1 - y_2}{\sqrt{\text{var}(y_1) + \text{var}(y_2)}}. \tag{5}$$

Illustrons-là maintenant avec la fonction d'Anscombe, la première, en 1948, à avoir été explicitement proposée à cette fin. Nous avons donc :

$$y_{\text{Ans}}(\text{Sans}) = \sin^{-1} \sqrt{\frac{9 + 3/8}{23 + 3/4}} \approx 0,6793;$$

$$y_{\text{Ans}}(\text{Avec}) = \sin^{-1} \sqrt{\frac{5 + 3/8}{26 + 3/4}} \approx 0,4648,$$

puis :

$$z(y_{\text{Sans}} - y_{\text{Avec}}) = \frac{0,6793 - 0,4648}{\sqrt{1/4 \times 23 + 2 + 1/4 \times 26 + 2}} \approx 1,514,$$

donnant lieu à une probabilité $P_{\text{Ans}} \approx 0,065$. Pour documentation, nous avons aussi $P_{\text{Cha}} \approx 0,065$, $P_{F-T/2} \approx 0,067$ et $P_{\text{Brute}} = 0,062$, cette dernière émanant du test très répandu du $z(p_1 - p_2)$ ($= 1,539$) et du khi-deux équivalent sur tableau 2×2 ($= 2,369 = 1,539^2$), ce dernier en mode déplié unilatéral.

Le lecteur aura peut-être remarqué, au dénominateur de la formule classique (5), que les variances de chaque groupe ne sont pas affectées d'une pondération (selon la

1. Le test classique (4) est par ailleurs l'approximation la plus sûre pour tester l'hypothèse « $\pi = 1/2$ », un hommage à Abraham de Moivre (1667-1754), qui créa la loi normale à partir d'une binomiale $B(n, \pi = 1/2)$ poussée vers l'asymptote.

2. L'objectif de cette correction appliquée à une approximation par une fonction continue (et évoquant une loi de probabilité continue comme la loi normale) est de rejoindre tout le poids de probabilité cumulé par la valeur x de la proportion x/n , c.-à-d., sur l'axe continu, le volume de probabilité compris entre $x - 1/2$ et $x + 1/2$ plutôt que le seul volume délimité dans l'intervalle $x - 1/2$ à x . Ainsi, la somme des probabilités binomiales discrètes cumulées jusqu'à x selon π et n correspondra (à peu près en général et très bien pour $\pi \rightarrow 1/2$) à l'intégrale normale de $(y - n \cdot \pi) / \sqrt{n \cdot \pi \cdot (1 - \pi)}$ pour y allant de $-\infty$ à $x + 1/2$.



taille du groupe), au contraire de la formule du test t , classique elle aussi, dans laquelle nous avons affaire à des estimateurs d'une hypothétique variance commune plutôt qu'à des valeurs « exactes ». D'après nos calculs (et nos graphiques), il appert que, dans ce cas-ci, la solution sans pondération est non seulement justifiée mais aussi plus proche de celle fournie par le calcul binomial exact.

Concluant sur notre exemple (fictif, bien sûr), nous constatons que les participants vaccinés sont deux fois moins nombreux à avoir subi l'agression de la grippe saisonnière, la proportion passant de 39% à 19%. Ces nombres, ajoutés à une probabilité de seulement 0,065 que le hasard soit imputable de cette différence, font que, jusqu'à preuve du contraire et la prudence étant de mise, les autorités de santé publique devraient inciter toutes les personnes admissibles à se faire vacciner, notamment les personnes âgées.

Note complémentaire-1. Dans le Matériel supplémentaire lié à cet article, le lecteur trouvera un tableau de concordance de huit tests de la différence entre deux proportions indépendantes : le test de LIDDELL (1978), le test de Fisher (SIEGEL & CASTELLAN, 1988), le test t , le test z , le test $z^{1/4}$ (avec réduction du numérateur de $1/4$ ($1/n_1 + 1/n_2$), le $z(y_{Ans})$ et le $z(y_{Cha})$, tous comparés au test « exact » de LAURENCELLE (2017b). Pour différentes combinaisons des tailles n_1 et n_2 et tous les cas pour lesquels $x_1/n_1 > x_2/n_2$, nous avons testé la différence $x_1/n_1 - x_2/n_2$ aux seuils α unilatéraux de 0,05 et 0,01. Le tableau indique les cas de sanctions de significativité discordantes entre chaque test versus le test exact. Comparativement au test exact : (1) le test de Liddell concorde presque parfaitement, étant basé sur un principe de calcul semblable; (2) le test de Fisher se montre indûment conservateur (et n'est pas approprié pour cette comparaison); (4) les tests t et z simples (sans correction), $z(y_{Ans})$ et $z(y_{Cha})$, aux comportements relativement semblables, sont un peu libéraux; (5) quant au test $z^{1/4}$, il rejoint le succès du test de Liddell pour les cas d'égalités des tailles, des discordances des deux types apparaissant dans les cas $n_1 \neq n_2$.

Note complémentaire-2. Le lecteur connaît aussi un test de différence entre deux moyennes jumelées, tels les tests en situations pré et post-traitement ou ceux comparant deux conditions d'expérience avec les mêmes participants, ce avec ou sans contrebalancement : un test de différence entre deux proportions jumelées en serait le cousin. Nous proposons plus loin une solution de calcul pour l'analyse de variance à un facteur avec mesures répétées, la méthode appliquée pouvant aussi servir au test de différence entre deux proportions.

L'analyse de variance de k proportions indépendantes

Du test t à l'analyse de variance, le chemin n'est pas long, la seconde étant simplement une extension du premier : seul le langage change. Il s'agit encore de déterminer si la variance entre les proportions comparées déborde significativement la variance imputable à leur incertitude échantillonnale, soit si $CM(\text{proportions}) > CM(\text{erreur})$ ou $F = CM_{\text{proportions}}/CM_{\text{intraproportion}} > 1$.

Soit les k proportions $p_j = x_j/n_j$, $j = 1$ à k , et leurs transformations $f(x_j, n_j)$. À l'instar de l'analyse classique, le terme $CM_{\text{proportions}}$, ou variance entre les proportions considérées, se calcule directement, soit $var(f(x_j, n_j))$, sans autre considération des données binaires brutes. Rappelons en effet qu'une proportion, ou sa transformation monotone, dénote déjà une moyenne, de sorte que la variance entre proportions en est une entre des moyennes de groupes. Pour compléter l'analyse, deux adaptations du calcul s'imposent :

1. La variance de chaque fonction $f(p)$ étant réputée constante, la variance dite « d'erreur », CM_{erreur} (ou $CM_{\text{intra-proportion}}$), dépend ici seulement de la taille n_j de chaque groupe, tel qu'indiqué aux formules (2a), (2b) ou (2b') et (2c). Par exemple, pour la fonction Chanter (2c), la variance réputée sera $1/(4n)$. Dans le cas de groupes à tailles n_j inégales, le calcul de variance basé sur la moyenne harmonique des tailles, soit $\sigma^2(\tilde{n}) = 1/(4\tilde{n})$ pour Chanter, convient, en utilisant $\tilde{n} = k/(1/n_1 + 1/n_2 + \dots + 1/n_k)$. Si les tailles ne sont pas trop inégales (p. ex. $n_{\text{max}} / n_{\text{min}} < 3$), la moyenne simple des variances d'erreur individuelles peut aussi convenir.³
2. La variance d'erreur de la proportion (et non pas de la donnée brute, comme c'est le cas en anova classique) étant ici réputée et calculée constante, ses degrés de liberté sont infinis, le quotient de variances F ayant alors $k - 1$ et ∞ degrés de liberté : la forme $(k - 1) \times F_{k-1, \infty}$ devient alors un khi-deux doté de $k - 1$ degrés de liberté.

Le test à faire est alors :

$$\begin{aligned} \text{khi-deux} &= (k - 1) \times CM_{\text{proportions}}/CM_{\text{erreur}} \\ &= (k - 1) \times var(f(p_j))/\sigma_e^2(\tilde{n}), \end{aligned} \quad (6)$$

en appliquant la formule de variance d'erreur appropriée à la fonction f : ce test est approximativement distribué comme χ^2 avec $k - 1$ degrés de liberté.

Exemple 3 (L'analyse de variance à un facteur sur 4 groupes indépendants). Comme premier exemple d'anova, nous avons opté pour une situation (formelle) comparant 4 groupes indépendants, de tailles modestes et

3. Le calcul simple selon \tilde{n} donne le même résultat que la moyenne des variances par la formule Chanter et une valeur légèrement plus basse par les formules Anscombe et Freeman-Tukey. Voir aussi (LAURENCELLE, 2017a) sur la pertinence de l'utilisation de la moyenne harmonique.



Tableau 1 ■ Sommaire des données et de l'analyse de variance de 4 proportions indépendantes

Groupe	<i>n</i>	<i>x</i>	<i>x/n</i>	Anscombe	Chanter	<i>F-T/2*</i>
1	12	6	0,5000	0,7854	0,7854	0,7854
2	8	3	0,3750	0,6701	0,6671	0,6726
3	15	13	0,8667	1,1720	1,1794	1,1662
4	10	4	0,4000	0,6918	0,6898	0,6936
<i>CM</i> _{proportions}			0,0517	0,0545	0,0567	0,0528
<i>CM</i> _{erreur}			0,0205	0,0224	0,0234	0,0224
Khi-deux			(7,569)	7,324	7,263	7,091

Note. * : La formule (2b') de la fonction Freeman-Tukey a été retenue pour fins de comparaison.

inégales; à titre d'illustration, nous en présentons le traitement via toutes les fonctions considérées ci-dessus. Le tableau 1 présente un sommaire des résultats et d'analyse.

La variance d'erreur globale (*CM*_{erreur}) rapportée au tableau 1 pour les trois fonctions de transformation est la moyenne simple des variances d'erreur $\sigma_e^2(n_j)$ associées aux proportions individuelles. Le calcul via la moyenne harmonique des tailles n_j et le simple recours à $\sigma_e^2(\bar{n})$ aurait donné 0,0224 pour les valeurs Anscombe et *F-T/2* : voir note 3 de bas de page.

Pour fins de comparaison, le tableau 1 présente aussi les calculs associés à une proportion brute, (*x/n*). Pour cette fonction, la variance d'erreur a une incertitude échantillonnale dépendant de la variable *x* et est donc évaluée selon ses degrés de liberté $n - 1$, par la formule $s_2(x/n) = p(1-p)/(n-1)$. Le *CM*_{erreur} rapporté est donc lui aussi une moyenne pondérée par ces mêmes degrés de liberté. Si on le jugeait convenable, le test classique approprié pour ce cas serait un test *F*, aux degrés de liberté $k - 1$ et $\sum n_j - k$, soit $F = 0,0517/0,0205 \approx 2,523$ versus $F_{3,41}$, la probabilité du *F* obtenu étant de 0,071.

La probabilité extrême associée aux khi-deux des trois fonctions, sous $\chi_{k-1}^2, \chi^2(k-1)$, est respectivement 0,062 (Ans), 0,064 (Cha) et 0,069 (*F-T/2*), signalant un résultat en bordure du seuil α de 0,05.⁴

La panoplie de tests complémentaires ordinairement associés à l'anova reste disponible pour le traitement des proportions via les fonctions angulaires : citons notamment les comparaisons de moyennes (= proportions) a posteriori et l'analyse polynomiale, la variance d'erreur exploitée référant toujours à un nombre infini de degrés de liberté. Pour illustration, l'application du test de l'étendue standardisée (Studentized range) de Tukey, dans notre exemple 3 comportant $k = 4$ 'moyennes', aurait pour

valeurs critiques 3,633 au seuil $\alpha = 0,05$ et 4,403 au seuil 0,01. Ainsi, par la fonction ANS, la différence la plus grande, celle entre les groupes 3 et 2, donnerait le calcul :⁵

$$q = \frac{1,1720 - 0,6701}{\sqrt{0,0223}} \approx 3,361,$$

un résultat qui n'atteint pas tout à fait le seuil de significativité de 0,05. Les cinq autres différences de moyennes possibles, parce que moins grandes, sont alors disqualifiées aussi.

Nous donnons un autre exemple (Exemple 5) des calculs ci-dessus dans la section suivante.

Exemple 4 (L'analyse de variance à devis factoriel sur 2 × 3 groupes indépendants). Pour second exemple d'une anova sur proportions indépendantes, prenons cette fois un devis à plan factoriel de forme A × B, que nous illustrons en exploitant la fonction Chanter. La compilation des données binaires brutes (fictives!) permet de soumettre le problème dans le tableau 2, la configuration des proportions sous la transformation Chanter (formule 2c) apparaissant à la figure 1.

L'analyse procède comme d'habitude, où nous devons produire des variances (ou CM, pour carré moyen) et tests *F* pour les sources de variation présentes, soit ici les facteurs A, B, respectivement de n_A et n_B niveaux, et leur interaction A × B, ce en tenant compte du poids relatif de chaque proportion concernée.

Calcul du CM(A). Chaque condition A est représentée par 3 moyennes en B, soit

$$\begin{aligned} moy(A_1) &= moy(A_1B_1, A_1B_2, A_1B_3) \\ &\approx 0,89230 \end{aligned}$$

et $moy(A_2) \approx 0,69430$. La variance de ces 2 deux proportions, égale à 0,019602, devient le CM de 0,05881, après

4. Par curiosité, nous avons effectué aussi le test classique du khi-deux d'interaction classique (sur matrice de fréquences 2 × 4) pour les données du tableau 1, obtenant khi-deux = 8,073 avec $p \approx 0,045$ sous χ_3^2 , et son valeureux compétiteur basé sur un calcul du maximum de vraisemblance, le test *G* (SOKAL & ROHLF, 1995), ici $G = 8,829$ avec $p = 0,032$ encore sous χ_3^2 . Ces deux tests conviennent mal au présent contexte, la comparaison des proportions y étant fortement influencée par la taille (n_j) de chaque groupe via la pondération des calculs.

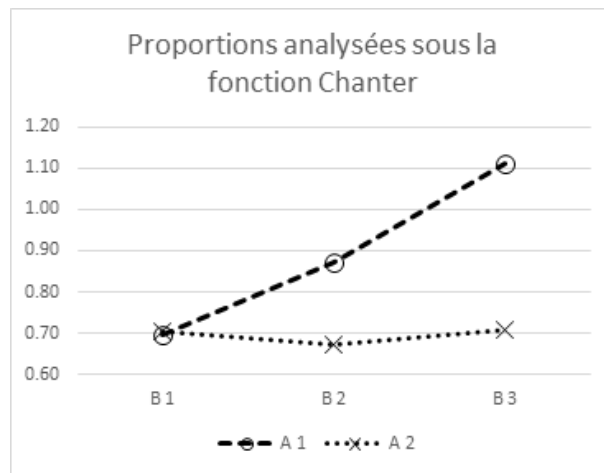
5. Nous utilisons ici le terme d'erreur global, celui appliqué à l'anova, ce par tradition et parce qu'il est basé sur plus d'information (et, le cas échéant, de degrés de liberté). L'utilisation des seuls termes d'erreur associés aux deux proportions comparées est elle aussi courante.



Tableau 2 ■ Ingrédients de l'analyse de variance de proportions selon un plan A × B

Groupe	n	x	p	y (Chanter)
A ₁ B ₁	22	9	0,406	0,6961
A ₁ B ₂	17	10	0,288	0,8715
A ₁ B ₃	26	21	0,808	1,1094
A ₂ B ₁	24	10	0,417	0,7034
A ₂ B ₂	26	10	0,385	0,6712
A ₂ B ₃	19	8	0,421	0,7082

FIGURE 1 ■ Configuration des proportions obtenues selon un devis A × B



pondération par sa 'charge' relative⁶ de 3, avec son degré de liberté de 2 - 1 = 1.

Calcul du CM(B). Même procédé pour traiter les 3 niveaux moyens du facteur B (B₁ ≈ 0,69976, B₂ ≈ 0,77135, B₃ ≈ 0,90880), chacun basé sur les deux niveaux de A. On obtient alors $CM(B) = 2 \times s^2(B_j) = 2 \times 0,011286 \approx 0,02257$, à degrés de liberté de 3 - 1 = 2.

Calcul du CM(A × B). Il existe différents procédés pour le calcul du terme d'interaction, le calcul par soustraction à partir de la variance totale du tableau AB, variance dotée de $n_A \times n_B - 1$ degrés de liberté, nous paraît le plus commode. Selon l'égalité formelle $var(AB_{totale}) = var(A) + var(B) + var(A \times B)$, nous calculons $CM(A \times B) = var(A \times B) = [(n_A \times n_B - 1) \times var(AB) - (n_A - 1) \times CM(A) - (n_B - 1) \times CM(B)] / [(n_A - 1)(n_B - 1)]$, chaque terme étant pondéré par ses degrés de liberté. Ce calcul donne ici $CM(A \times B) = (5 \times 0,02913 - 1 \times 0,05881 - 2 \times 0,02257) / 2$, avec ses degrés de liberté de $(2-1)(3-1) = 2$.

Calcul du terme CM(erreur). La variance d'erreur de la fonction Chanter est 'stabilisée' comme celle des fonctions Anscombe et Freeman-Tukey; il reste qu'il faut y considérer les tailles n_j rencontrées. Nous optons cette fois-ci de recourir à la moyenne harmonique des tailles, une solution qu'on applique aussi couramment en anova classique (LAURENCELLE, 2017a). Utilisant les 6 n_j , nous obtenons $\tilde{n} \approx 21,7786$, d'où $CM(erreur) = \sigma_e^2(\tilde{n}) = 1 / (4 \times \tilde{n}) \approx 0,01148$.

Calcul des tests khi-deux. Toutes les sources étudiées (A, B, A × B) sont testées avec le khi-deux obtenu par le quotient du CM approprié pondéré par ses degrés de liberté et divisé par le CM d'erreur. Le tableau 3 présente le sommaire des résultats du calcul.

De la lecture du tableau 3, et nonobstant l'interaction apparente qui suggère un effet B plus grand pour la condition A₁, le seul effet qui ressort significatif ($p \approx 0,024$) tient à la différence moyenne de niveau entre A₁ et A₂.

Tous les plans d'analyse utilisant l'une ou l'autre

6. Noter que l'ingrédient du calcul - une proportion basée sur n unités - n'est pas la donnée brute elle-même, de sorte que les tailles n_j n'apparaissent pas dans les calculs de l'anova, sinon dans le calcul du CM d'erreur, lequel est obtenu par pondération des tailles échantillonales concernées. Noter que les calculs rapportés tiennent compte de plus de décimales que celles présentées ici, ce qui explique certaines discordances possibles à la dernière décimale imprimée.



Tableau 3 ■ Sommaire de l'analyse de variance des données présentées au Tableau 2.

Source de variation	Degrés de liberté	Carré moyen	Khi-deux	Probabilité
Totale	5	0.0291		
A	1	0.0588	5.1231	0,024
B	2	0.0226	3,9326	0,140
A × B	2	0.0209	3.6346	0.162
Erreur	∞	0.0115		

des fonctions transformatrices illustrées ici peuvent être traités de manière analogue, incluant les plans simples avec mesures répétées, dont le traitement spécifique est explicité plus bas, ainsi que les plans complexes, tels les plans à trois facteurs croisés ou plus, ceux à facteurs emboîtés, ceux à interactions confondues, et cætera. La seule différence essentielle par rapport à l'anova sur des données habituelles à référence normale tient au terme d'erreur, lequel reste toujours constant et le même à tous les étages de l'analyse.

L'analyse de variance de k proportions jumelées (« à mesures répétées »)

Il est souvent intéressant de comparer le taux de réponse, le niveau de succès des mêmes personnes ou objets d'étude après qu'ils aient été soumis à deux ou quelques conditions d'expérimentation différentes : c'est ici la situation classiquement dévolue à l'analyse de variance avec mesures répétées, le test t sur deux moyennes jumelées en étant un cas particulier. Deux caractéristiques distinguent cette forme d'analyse : pour la première, le nombre de mesures par condition étant égal au nombre de sources ou personnes, il est donc égal dans toutes les conditions, tandis que, pour la seconde caractéristique, étant donné une certaine stabilité du niveau de réponse produite par chaque source échantillonnale, les différences constatées entre les moyennes de conditions sont de

généralité plus sûre, telle que reflétée par un terme d'incertitude (le CM_{erreur}) réduit.⁷ Sous des conditions ordinaires, l'anova avec mesures répétées est donc plus puissante que celle appliquée à des données de sources échantillonnales toutes indépendantes.

La détermination du CM_{erreur} . Le groupe, l'échantillon évalué d'une condition à l'autre est constitué de sources (personnes ou objets) caractérisées chacune par un niveau de mesure typique, lequel niveau varie ordinairement d'une source à l'autre. Pour chaque source donnée, la mesure obtenue sous chaque condition et, a fortiori, la moyenne de ses mesures dans les différentes conditions imposées tendront à refléter ce niveau typique individuel, le tout étant plus ou moins brouillé par des variations inexplicables. La stabilité relative de réponse de chaque source va créer de la corrélation positive dans le tableau des résultats : à la limite, pour une stabilité parfaite du niveau individuel, toutes les sources produiront exactement le même profil de réponse, ce à différentes hauteurs de mesure proportionnelles aux dits niveaux typiques.

Dans l'approche classique pour ce cas, la variance d'incertitude sur la comparaison entre les moyennes de conditions ou moments de mesure correspond à la mesure d'interaction, analogiquement à la variance d'interaction $s^2(A \times B)$ traitée plus haut. Pour un devis d'analyse à un facteur (ou dimension), cette mesure est habituellement obtenue par soustraction, selon :

$$CM_{\text{erreur}} = \frac{k(n-1)var(totale) - (n-1)var(sources) - (k-1)var(conditions)}{(n-1)(k-1)}, \tag{7}$$

le carré moyen résultant ayant $(n-1)(k-1)$ degrés de liberté. Une autre voie de calcul, relativement plus simple, et applicable à l'analyse des proportions, procède par la for-

mule suivante :

$$CM_{\text{erreur}} = \frac{knS - V}{k(k-1)} \tag{8}$$

où S dénote la somme des variances des k moyennes

7. Ce terme « d'erreur » devient alors un terme d'interaction Condition × Source, dont a été expurgée une part (souvent importante) de la variabilité inter-individuelle des sources.



de conditions (ici, les k proportions simples) et V , la variance de la somme des mesures obtenue par chacune des n sources (ici, le x_i de chaque proportion). La portion de réduction du CM_{erreur} depuis sa valeur initiale (sans corrélation) de S/k est exactement égale à l'estimé pondéré « α_1 », ou α de CRONBACH (1951) du tableau de données, après unitarisation, soit $\alpha_1 = \alpha / (k - (k - 1) \cdot \alpha)$, le coefficient α_1 reflétant l'intercorrélacion moyenne entre les conditions.⁸ Ce coefficient peut s'obtenir ici par :

$$\alpha_1 = \frac{V - n \cdot S}{(k - 1) \cdot n \cdot S} \tag{9}$$

Par similitude de structure entre l'anova classique et celle développée ici pour les proportions, nous proposons le calcul suivant :

$$CM_{\text{erreur}} = \sigma_e^2(n)(1 - \alpha_1), \tag{10}$$

la variance d'erreur applicable étant celle de la fonction angulaire utilisée.

Exemple 5. (L'analyse de variance avec mesures répétées sur 5 moments successifs). Le tableau 4 présente les données brutes et quelques ingrédients utiles aux calculs de l'anova à mesures répétées sur 5 moments ou conditions. Noter qu'ici, comme c'est le cas de l'anova classique, les données brutes (individuelles) sont requises pour établir le niveau de corrélation qu'elles présentent d'une mesure à l'autre. Aussi, et comme les tailles n_j des proportions sont égales, toutes les fonctions de transformation conviennent, s'équivalant à peu près, et nous proposons ici la fonction (2a) d'Anscombe.

Comme il s'agit ici d'une anova de structure pareille à celle de notre exemple 3, soit celle d'un tableau à facteur unique à k niveaux, les calculs sont les mêmes, jusqu'à la finalisation du CM_{erreur} , lequel est influencé par la corrélation probable des k mesures issues de chacune des n sources. Cette corrélation est reflétée par le coefficient α_1 (éq. 8).

Nous avons ici $k = 5$ et $S = 0,0126 + 0,0227 + \dots + 0,7500 \approx 0,09722$. Pour V , les totaux de chacune des $n = 12$ sources apparaissant au tableau ont pour variance $\approx 2,9697$. Quant à la variance d'erreur attachée aux proportions transformées, elle est constante et dépend uniquement de leur taille commune n et du degré de corrélation de la matrice des données tel qu'estimé par le coefficient α_1 . Par (9), nous obtenons enfin $\alpha_1 \approx (2,9697 - 12 \times 0,09722) / [(5 - 1) \times 12 \times 0,09722] \approx 0,3863$,⁹ et le CM_{erreur} s'obtient par la formule (10).

8. Un lien intéressant est à faire avec le traitement des tableaux de réponse (de format binaire ou non) rencontrés en psychométrie, via l'indice dit de consistance interne, le α de Cronbach (1951; voir aussi Laurencelle, 1998, obtenu par $k / (k - 1) \times (1 - nS/V)$, empruntant la notation de l'équation (9). Le CM_{erreur} cherché ici est égal à $nS/k(1 - \alpha_1)$, nS/k correspondant au CM intra-conditions, dans le langage de l'anova. Noter que l'indice α_1 est un estimé de la corrélation moyenne entre conditions ou moments de mesure, estimé dûment influencé par la fluctuation des (paires de) variances.

9. Tel que documenté dans la note 8 de bas de page, le lecteur pour aussi utiliser le fameux coefficient α de CRONBACH (1951), pour le transfert duquel vers α_1 est clairement exposé. Noter que, pour des conditions à variances toutes égales, la valeur du α_1 devient identique à celle de l'intercorrélacion moyenne de toutes les conditions.

Les calculs finaux pour ce devis plutôt simple sont :

$$\begin{aligned} CM_{\text{conditions}} &= \text{var}(y_j) = 0,04381, \\ CM_{\text{erreur}} &= \sigma_e^2 y_{\text{ANS}}(n) \times (1 - \alpha_1) \\ &= 1 / (4 \times 12 + 2) \times (1 - 0,3863) \\ &\approx 0,01227, \end{aligned}$$

pour un khi-deux :

$$\begin{aligned} \text{khi-deux} &= (k - 1)CM_{\text{conditions}} / CM_{\text{erreur}} \\ &= 4 \times 0,04381 / 0,01227 \approx 14,278, \end{aligned}$$

une valeur de probabilité droite de 0,006 au χ_4^2 .

L'anova des proportions dans un devis multifactoriel avec mesures répétées procéderait de même, en appliquant les adaptations idoines, adaptations qui incluraient notamment l'estimation d'un coefficient α_1 global par la moyenne des coefficients attenants à chaque combinaison des facteurs soumis à la répétition de mesure.

Test d'une hypothèse globale complémentaire. En complément de notre exemple 5, considérons que les devis à facteur unique avec mesures répétées mettent souvent en jeu des conditions d'expérimentation graduées (avec ou sans contre-balancement de l'ordre séquentiel de ces conditions entre les participants), conditions pour lesquelles le chercheur est intéressé à déterminer s'il y a progression de la réponse d'une condition à l'autre par des valeurs montantes ou bien descendantes de la mesure. Supposons que notre exemple 5 soit de ce type, et que le chercheur veuille déterminer si la contre-hypothèse de variation croissante des proportions est crédible, de la condition (ou moment) 1 à la condition (ou moment) 5. Le test de variation monotone (BARLOW, BARTHOLOMEW, BREMNER & BRUNK, 1972; LAURENCELLE, 1993; LAURENCELLE & DUPUIS, 2002a) répond précisément à ce besoin.

Les données concernées, ici $k = 5$ proportions y_{ANS} , sont reproduites à la première rangée du tableau 5. Pour faire court et pour ce cas particulier, il s'agit d'abord de régulariser la série de valeurs observées de sorte que la variation monotone, c.-à-d. de même orientation numérique, ne soit pas contredite; cela s'opère en 'amalgamant' des valeurs contiguës de manière à assurer une variance maximale de la série produite. Ici, la combinaison des données des conditions 2 et 3 suffit, donnant 0,7460 pour chacune des deux. De $r = k = 5$ valeurs originales, on est passé à $r = k - 1 = 4$ valeurs distinctes qui respectent maintenant



Tableau 4 ■ Données binaires brutes et ingrédients de calcul pour l’analyse de variance de 5 proportions jumelées

Source	Moment 1	Moment 2	Moment 3	Moment 4	Moment 5	Total
1	0	1	0	1	1	3
2	0	1	0	0	0	1
3	0	0	1	0	1	2
4	1	1	1	1	1	5
5	0	0	1	0	1	2
6	1	1	1	1	1	5
7	0	0	0	0	1	1
8	0	0	0	0	0	0
9	0	1	0	1	1	3
10	0	0	0	0	0	0
11	0	1	1	1	1	4
12	0	0	0	1	1	2
<i>x</i>	2	6	5	6	9	
<i>p</i>	0,1667	0,5000	0,4167	0,5000	0,7500	
<i>var(p)*</i>	0,0126	0,0227	0,0221	0,0227	0,0170	
<i>y</i> _{ANS}	0,4463	0,7854	0,7066	0,7854	1,0304	

Note. $var(p) = p(1 - p)/(n - 1)$

Tableau 5 ■ Proportions transformées *y*_{Ans} observées et leurs valeurs régularisées

	1	2	3	4	5
Observées (<i>y</i>)	0,4463	0,7854	0,7066	0,7854	1,0304
Régularisées (<i>y</i> _R)	0,4463	0,7460	0,7460	0,7854	1,0304

l’ordre croissant. La statistique à faire, notée \bar{X}^2 , est alors :

$$\bar{X}^2 = (r - 1)CM_{y_R} / CM_{\text{erreur}} \quad (11)$$

Le carré moyen décrivant les données régularisées *y*_R associées aux conditions est ici $CM_{y_R} \approx var(y_R) = 0,04303$ (le CM_y original étant 0,04381), le test \bar{X}^2 est donc égal à $(4 - 1) \times 0,04303 / 0,01227 \approx 10,520$. La statistique \bar{X}^2 correspond à un composé de variables χ^2 pondérées : LAURENCELLE et DUPUIS (2002b, 2002a) esquissent la théorie et donnent des tables complètes. La valeur critique pour $k = 5$ au quantile 0,975 est de 9,237, ce qui autorise à conclure à une progression des proportions significative à $p < 0,025$.¹⁰ ABELSON et TUKEY (2013) proposent une autre statistique, approximative celle-là, pour vérifier l’hypothèse de variation monotone, solution qu’on peut aussi adapter au traitement des proportions.

Comparaison avec l’anova classique. Pour conclure sur notre exemple 5, que serait-il sorti d’une analyse classique des données du Tableau 4 ? Le lecteur vérifiera facilement que, pour ces données, par une méthode ou une autre,

nous obtenons :

$$CM_{\text{conditions}} = n \times var(p_j) \approx 0,5250,$$

$$dl = k - 1 = 4,$$

$$CM_{\text{erreur}} = (knS) / [k(k - 1)] \approx 0,1432,$$

$$dl = (k - 1)(n - 1) = 44,$$

puis :

$$F = CM_{\text{conditions}} / CM_{\text{erreur}} \approx 0,5250 / 0,1432 \approx 3,666.$$

Sous la loi *F* dotée de 4 et 44 degrés de liberté, ce quotient rend une probabilité de 0,012, versus la probabilité de 0,006 obtenue plus légitimement ci-dessus au moyen de la transformation *y*_{ANS}.

Le test *z* sur deux proportions corrélées et la solution McNemar

Le test *t* étant un cas particulier de l’analyse de variance, on peut effectuer le test de différence de deux proportions par la méthode proposée pour un nombre $k (\geq 2)$ quelconque de moyennes à comparer. Nous offrons tout de même la formule plus simple via le test *z*, en prolongement

10. La documentation indiquée pour le test de variation monotone couvre aussi les analyses de variance classique, avec degrés de liberté finis du terme d’erreur et valeurs critiques de la statistique appropriée \bar{E}^2 .



de celui de la formule (5), soit :

$$z = \frac{y_1 - y_2}{\sqrt{2 \text{var}(y)(1 - r_{1,2})}} \tag{12}$$

Dans cette formule, le y_j dénote la valeur de chaque proportion transformée selon le choix de l'utilisateur, $\text{var}(y)$ la variance correspondante (et égale pour les deux proportions, les tailles étant égales), et $r_{1,2}$, la corrélation simple entre les deux séries binaires comparées.¹¹

Considérant l'Exemple 5 plus haut et les données des Moments 1 et 2 du Tableau 4, y a-t-il un écart significatif entre ces deux conditions? En recourant cette fois aux formules (2b) de Freeman et Tukey (pour illustration), nous obtenons :

$$y_1 \approx 1,1047; y_2 \approx 1,5708; \\ \text{var}(y) \approx 0,0800; r_{1,2} \approx 0,4472,$$

d'où :

$$z = \frac{0.9042 - 1.5708}{\sqrt{2 \frac{1}{12+0.5}(1 - 0.4472)}} \approx -2,242,$$

un résultat de probabilité unilatérale de 0,012 ou de 0,025 en mode bilatéral.

Versus le test de McNemar. L'approche de McNEMAR (1947) pour la comparaison de deux proportions corrélées et sa célèbre procédure de test sont tout différentes. Elles se basent sur le raisonnement suivant. Les sources (ou personnes) qui gardent inchangée leur réponse ne nous apprennent rien sur l'orientation probable d'un changement dans la population puisqu'elles ne changent pas, l'information sur le changement émanant plutôt des sources dont la réponse s'est modifiée. Pour celles-là, y a-t-il une tendance nette à passer du Oui au Non, du Contre au Pour, du Malade au Guéri, etc. ? Pour illustration, utilisons une nouvelle fois les données de l'exemple 5 dont nous réorganisons les totaux selon une matrice 2 × 2 (voir Tableau 6) comme le fait McNemar.¹²

L'examen de la matrice montre que 8 sources (6 «00» + 2 «11») ont fourni la même réponse aux Moments 1 et 2 tandis que les 4 sources qui changent l'ont fait en passant toutes du «0» au «1», ces changements faisant évoluer la proportion de «1» de 6/12 à 10/12, pour une différence de $\Delta = p_1 - p_2 \approx 0,333$. Sur la seule base des sources qui changent, peut-on conclure à une tendance ?

Le test de McNemar, qui vérifie si la part de changements positifs diffère significativement de celle des

négatifs, repose sur l'hypothèse nulle stipulant que, le changement se faisant au hasard, la probabilité de changer dans chaque direction est de 1/2, d'où la pertinence d'un test binomial facile à réaliser. Ce test peut se faire par un calcul exact (ou via la consultation de tables binomiales), ou par l'excellente formule d'approximation suivante :¹³

$$z = \frac{|n_{01} - n_{10}| - 1}{\sqrt{n_{01} + n_{10}}} \tag{13}$$

où n_{01} et n_{10} désignent le nombre de conversions de réponse dans un sens et dans l'autre. Dans le cas présent, avec $n_{01} = 4$ et $n_{10} = 0$, nous obtenons facilement $z = 1,500$, de probabilité normale (unilatérale) de $\approx 0,067$, la probabilité binomiale exacte étant ici $1/2^4 = 0,0625$. Selon un seuil de significativité de 0,05 en mode bilatéral, la probabilité unilatérale limite eût été de 0,025, laquelle n'est pas atteinte dans notre exemple, tandis que, d'autre part, la différence de proportions, $\Delta = 0,333$ avec $p \approx 0,012$, l'était selon notre formule (11).

Le test de différence entre proportions corrélées des Moments 1 et 2, tel que concrétisé par la formule (11), nous indique, s'il est significatif, qu'il y aurait tendance à ce que la proportion de «1» augmente (ou diminue) dans la population si celle-ci était mesurée pour chaque source d'une condition (ou moment) à l'autre. Quant au test de McNemar, si significatif, il nous indique que les éléments de la population qui changeraient leur réponse le feraient préférentiellement dans une direction particulière. Il s'agit donc d'un test de la polarité des changements, test distinct de l'autre, assis sur une hypothèse nulle différente et orienté vers une conclusion autre. Le tableau 2 × 2 précédent et les deux tableaux suivants en font voir le contraste.

A	0	1	B	0	1
0	60	4	0	100	6
1	0	20	1	0	100

Au tableau A représentant 84 sources, la sanction du test de McNemar est la même qu'au tableau 5, étant encore basée sur 0 et 4 changements opposés et à probabilité $(1/2)^4 \approx 0,0625$. Quant au test de différence entre deux proportions ($\Delta = -0,048$), utilisant la transformation y_{F-T} , nous obtenons $z = -2,037$ de probabilité (unilatérale) 0,021, significative encore, nonobstant une différence de proportions drastiquement réduite. Au tableau B avec 206 sources, le contraste de 0 à 6 entre les

11. Cette corrélation peut être obtenue par un calcul semblable à celui du coefficient α_1 (8), soit $r_{1,2} = (V - nS)/(2ns_1s_2)$, où V est la variance de la somme des deux données des répondants, S la somme des variances des deux proportions, s_1 et s_2 les écarts-types correspondants. Elle correspond ici au coefficient α_1 (8) défini plus haut (parce qu'elle n'est basée ici que sur une seule covariance).

12. La matrice de McNemar se présente plutôt avec les rangées interverties, les «1» étant placés en première rangée.

13. McNEMAR (1947) propose plutôt un khi-deux de formule $(n_{01} - n_{10})^2/(n_{01} + n_{10})$, équivalant à $z = (n_{01} - n_{10})/\sqrt{n_{01} + n_{10}}$, omettant la correction de -1 au numérateur préconisée par plusieurs auteurs et vérifiable par comparaison avec la valeur exacte. Le calcul de McNemar donne $z = -2,000$ de probabilité $\approx 0,023$.



Tableau 6 ■ Réarrangements des données des Moments 1 et 2 du Tableau 4

Moment 1	Moment 2		Total
	0	1	
0	6	4	10
1	0	2	2
Total	6	6	12

changements donne pour McNemar une probabilité de $(1/2)^6 \approx 0,016$, significative cette fois, tout comme l'est la différence entre proportions encore réduite ($\Delta = -0,029$) selon la formule (11), avec $z \approx 2,476$ et $p \approx 0,007$. Ainsi, étant centré sur les seuls changements de réponse, le test de McNemar reste insensible à la configuration générale des données¹⁴ et sa conclusion s'applique principalement, sinon uniquement, à ceux-là.

LU (2010), qui reconnaît la lacune du test de McNemar et revoit quelques variantes suggérées dans la littérature, propose à son tour d'inclure les réponses non changeantes dans un nouveau test mieux contextualisé, ce nouveau test concernant encore les sources changeantes.¹⁵

Selon notre analyse, un test significatif de McNemar ou d'une de ses variantes nous informe du degré probable de polarisation du changement (binaire) pour les éléments de population qui changent, une polarisation importante pouvant advenir dans une fraction minimale de la population considérée. Quant à prédire le changement de proportions dans la population, nous concluons donc que le test de McNemar doit être employé avec prudence et à bon escient, et qu'il ne constitue pas une procédure générale adéquate pour comparer deux proportions corrélées.

Récapitulation et conclusions

La proportion $p = x/n$, régie par la loi binomiale $B(x|\pi, n)$ et d'espérance $\mu = n \cdot \pi$, lui doit une variance liée à l'espérance, soit $\sigma^2 = \mu \cdot (1 - \pi)$, de même qu'une non-normalité marquée (asymétrie $\gamma_1 = (1 - 2\pi)/\sigma$; aplatissement $\gamma_2 = [1 - 6\pi(1 - \pi)]/\sigma^2$), laquelle se résorbe à $\pi = 1/2$ dans une distribution symétrique et un aplatissement de $-2/n$. Ces deux caractéristiques contrarient les conditions prescrites pour l'utilisation de tests à approximation normale, un constat qui a amené les statisticiens à chercher des fonctions mathématiques supplétives dont le but principal est de ramener la variable « proportion » dans le corpus des variables traitables par les analyses statistiques à base de loi normale. Les trois fonctions (dénotées

y) étudiées ici, celles d'Anscombe (1948 – Ans), de Freeman et Tukey (1950 – F-T) et de Chanter (1975 – Cha) ont, en espérance, une variance indépendante de la moyenne et constante par rapport à π , et leur non-normalité, encore présente, est atténuée par rapport à celle de p et, dans plusieurs cas, fortement atténuée par la transformation. Que doit-on en penser ?

Notre « Étude comparative » LAURENCELLE (2021a) du comportement des trois fonctions « y » retenues, incluant des données complémentaires basées sur le calcul binomial exact et sur les « moins bonnes » approximations normales existantes, a permis de documenter les propriétés des fonctions lorsque exposées à différents contextes paramétriques et pratiques. Quel est l'effet du paramètre binomial π sur le biais des fonctions, leur variance, leur normalité, leur respect du seuil de significativité α , la puissance statistique qui s'y attache ? Quels sont les effets de la taille n , des diverses combinaisons de π et de n , de l'inégalité de tailles des proportions comparées ? Peut-on faire l'analyse de variance des proportions, et qu'en est-il de la comparaison de proportions corrélées ? L'examen de la documentation jointe, documentation que nous savons incomplète, nous a conduit à quelques conclusions provisoires. Comme il pourra en juger lui-même, le dossier accompagnateur de « preuves » (ibid.) ou, plutôt, d'illustrations, nous a paru d'une assez grande complexité. C'est donc en toute humilité que nous soumettons au lecteur nos conclusions.

1. Le test sur une proportion, répondant à l'hypothèse nulle « $\epsilon(x/n) = \pi$ », a sa solution simple par le calcul binomial exact ou, pour $\pi = 1/2$, par l'excellente approximation normale valide même pour de petites tailles n . Pour des tailles élevées et $\pi \neq 1/2$, les fonctions « y » concordent généralement avec le calcul binomial et sont plus exactes que ne l'est l'approximation normale classique.

2. Variances rendues moins discordantes par la transformation angulaire et non-normalité atténuée font que le test de différence entre deux proportions via les fonc-

14. Ce paradoxe apparent confine à la distinction entre grandeur d'effet et significativité. Le lecteur aura noté que, en comparant les données du Tableau 5 et celles de notre tableau A, les tests ($z = -2,242$ et $2,037$) sont comparables et tous deux significatifs bien que la différence des proportions (0,333 vs 0,048) ait dramatiquement diminué, cette réduction d'écart étant compensée pour un indice de cohérence plus grande des données (80 sources sur 84 ayant répété leur première réponse), générant une corrélation plus forte (de $r \approx 0,447$ à $r = 0,884$). Grandeur d'effet effectivement réduite au niveau des proportions, certes, mais généralisabilité meilleure (« Il y a une tendance significative à une élévation de la proportion d'une condition à l'autre dans la population »), ce qui respecte exactement le concept de significativité statistique.

15. Une formule du test de LU (2010, p. 3352) est $z = (n_{10} + n_{01}) / \sqrt{(n_{10} + n_{01})[2 - (n_{10} + n_{01})/N]}$.



tions « y » s'en tire avec les honneurs, primant souvent en exactitude à la fois sur le test $z(p_1 - p_2)$ classique et son équivalent de forme khi-deux. L'inégalité des tailles n_1 et n_2 affecte l'exactitude des tests exploitant les fonctions « y » tout comme du test z classique et du calcul exact (réalisé selon la solution Liddell 1978 ou la solution Laurencelle 2005, 2017b, 2021b).

3. Les fonctions « y » rendent possible l'analyse de variance, une méthode qu'il serait déraisonnable d'appliquer avec les proportions p simples. L'anova exige de légers ajustements de calcul et a notamment recours à la loi χ^2 plutôt qu'au F de Fisher.

4. Pour tester la différence entre deux ou plusieurs proportions corrélées, nous proposons un procédé de calcul équivalent au calcul de l'anova avec mesures répétées et basé comme lui sur l'intercorrélation entre les groupes de données comparées. L'intercorrélation est estimée directement à partir des données binaires brutes. Le procédé proposé met en lumière la limite du test de McNemar (1947; Siegel et Castellan, 1988; Sokal et Rohlf, 1995), lequel statue spécifiquement et seulement sur la polarité de changement de réponse des sources échantillonales qui changent et non sur une tendance générale de la population échantillonnée.

5. La puissance statistique du test sur une proportion ou sur la différence entre proportions évolue de façon comparable entre les trois fonctions y étudiées (LAURENCELLE, 2021a), cela en émulation avec le calcul binomial exact, ce dernier présentant un avantage léger mais consistant. Dans le cas de la différence entre proportions de tailles n_j inégales, certaines perturbations (notamment de l'exactitude) apparaissent mais restent inconsistantes.

6. Les fonctions d'Anscombe, de Chanter et de Freeman et Tukey sont-elles d'intérêt égal? Nous devons formuler ici une réponse nuancée et tenant compte du contexte d'application.

a. Pour la plupart des applications, telles la différence entre deux ou plusieurs proportions, la différence entre des proportions corrélées et l'élaboration de composés linéaires de proportions, les trois fonctions ont un avantage appréciable, voire unique, sur le traitement des valeurs brutes des proportions.

b. Des trois fonctions angulaires considérées, celle de F-T s'est révélée souvent moins efficace, alors que Ans et Cha prenaient habituellement la tête, la fonction Cha présentant une légère prédominance.

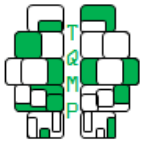
c. Les trois fonctions ont un biais (mathématique) qui dépend du paramètre π et s'estompe progressivement pour une taille n croissante, biais d'interaction donc et qui peut avoir un effet pervers sur la comparaison entre des proportions (écartées de $1/2$) calculées sur des tailles différentes. Seule la fonction Cha s'attaque de front à ce

problème et parvient à en atténuer substantiellement les effets. Partant, nous recommandons fortement l'usage de cette fonction pour les contextes à petits échantillons ($n < 100$) mettant en jeu des proportions p telles que $|p - \pi| > 0,15$.

À titre de 7^e et dernière considération, nous reconnaissons que plusieurs des procédures élaborées et illustrées ici sont exploratoires et que leur utilisation pratique repose sur une seule présomption de validité. Nous offrons donc au lecteur intéressé l'option d'asseoir ces procédures sur des bases plus complètes et d'y apporter corrections, améliorations et extensions.

Références

- ABELSON, R. P. & TUKEY, J. W. (2013). Efficient utilization of non-numerical information in quantitative analysis general theory and the case of simple order. *Annals of Mathematical Statistics*, 34, 1347-1369.
- ANCELLE, T. (2017). *Statistique, épidémiologie (4e édition)*. Paris : Éditions Maloine.
- ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35, 246-254.
- BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. & BRUNK, H. D. (1972). *Statistical inference under order restrictions*. New York : Wiley.
- BARTLETT, M. S. (1947). The use of transformations. *Biometrics*, 3, 39-52.
- CHANTER, D. O. (1975). Modifications of the angular transformation. *Journal of the Royal Statistical Society Series B (Applied Statistics)*, 24, 354-359.
- CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- FREEMAN, M. F. & TUKEY, J. W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, 21, 607-611.
- HOWELL, D. C., YZERBYT, V. & BESTGEN, Y. (2008). *Méthodes statistiques en sciences humaines (2e édition)*. Bruxelles : De Boeck.
- LAURENCELLE, L. (1993). Deux tests de variation monotone pour l'analyse de variance. *Lettres Statistiques*, 9, 69-91.
- LAURENCELLE, L. (1998). *Théorie et techniques de la mesure instrumentale*. Québec : Presses de l'Université du Québec.
- LAURENCELLE, L. (2005). Le test binomial exact de la différence entre deux proportions. *Lettres statistiques*, 12, 1-13.
- LAURENCELLE, L. (2017a). L'analyse de variance pour des groupes inégaux et la solution du n harmonique : une question d'équité [The unweighted "harmonic mean" solution for unbalanced anova designs : A detailed ar-



- gument]. *The Quantitative Methods for Psychology*, 13, 95-104.
- LAURENCELLE, L. (2017b). 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, 166-198.
- LAURENCELLE, L. (2021b). The exact binomial test between two independent proportions : A companion. *The Quantitative Methods for Psychology*, 17, 76-79.
- LAURENCELLE, L. & DUPUIS, F. (2002a). *Statistical tables, explained and applied*. Singapour : World Scientific.
- LAURENCELLE, L. & DUPUIS, F. (2002b). *Tables statistiques expliquées et appliquées*. Québec : Le Griffon d'argile.
- LIDDELL, D. (1978). Practical tests of 2x2 contingency tables. *The Statistician*, 25, 295-304.
- LU, Y. (2010). A revised version of McNemar's test for paired binary data. *Communications in Statistics*, 39, 3525-3539.
- MCNEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153-157.
- SIEGEL, S. & CASTELLAN, N. J. (1988). *Nonparametric statistics (2e édition)*. New York : McGraw-Hill.
- SOKAL, R. R. & ROHLF, F. J. (1995). *Biometry : the principles and practice of statistics in biological research (3e édition)*. New York : W.H. Freeman.

Open practices

- 📄 The *Open Material* badge was earned because supplementary material(s) are available on [the journal's web site](#).

Citation

- LAURENCELLE, L. (2021c). 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(3), 272-285. doi :[10.20982/tqmp.17.3.p272](https://doi.org/10.20982/tqmp.17.3.p272)

Copyright © 2021, 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: 07/06/2021 ~ Accepted: 20/02/2021