



# Le test binomial exact de la différence entre deux proportions et ses approximations

## The exact binomial test between two proportions and its approximations

Louis Laurencelle<sup>a</sup> 

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


**Abstract** ■ Though it is a commonplace and quite frequent operation, the comparison of two independent proportions remains a problematic and ill understood issue in textbooks on applied statistics and for researchers. On the one hand, the comparison may be arrayed as a  $2 \times 2$  contingency table and be referred to chi-square type calculations; on the other hand, it can be viewed as the difference between means based on 0 / 1 data, this perspective leading to alternative normal-type treatments. The main known solutions for deciding upon the difference between two proportions are reviewed and briefly discussed. An original, theoretically deduced solution that we deem ‘exact’ is then proposed. We conclude with a short numerical study that endorses two well-known approximate solutions.

**Résumé** ■ Situation statistique banale, voire quotidiennement rencontrée, la comparaison de deux proportions indépendantes demeure une opération problématique dans les ouvrages de statistique appliquée comme dans la pratique. D’un côté, cette comparaison peut être présentée dans le format d’un tableau de contingence  $2 \times 2$ , ce qui l’associe aux statistiques habituelles d’un tel tableau; d’un autre côté, la différence entre deux proportions équivaut à la différence entre deux moyennes indépendantes, en considérant que les valeurs moyennées sont des « 0 » et des « 1 », ce cas admettant un autre type de traitement. Nous revoyons ici les principales solutions proposées pour l’étude de la différence entre deux proportions, et nous élaborons par déduction notre propre solution, que nous croyons exacte. Dans une courte étude numérique, nous établissons enfin les vertus de deux approximations déjà connues.

**Keywords** ■ inference between two proportions, Exact binomial test, Fisher’s exact probability test, Binomial-to-normal approximation, Continuity correction,  $2 \times 2$  contingency table..

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

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

 LL: 0000-0003-3448-2872

 [10.20982/tqmp.13.2.p141](https://doi.org/10.20982/tqmp.13.2.p141)

### Introduction

Le premier échantillon, de  $n_1$  éléments, contient  $x_1$  cas favorables et  $r_1 = n_1 - x_1$  cas non favorables; de même pour le second échantillon, avec  $n_2$ ,  $x_2$  et  $r_2$ , selon le tableau 1.

Les proportions de cas favorables observées dans les échantillons sont  $p_1 = x_1/n_1$  et  $p_2 = x_2/n_2$ . La question posée est la suivante:

*Au seuil de signification  $\alpha$  ( $0 < \alpha < 1/2$ ), est-il plausible que ces deux échantillons proviennent d’une même population (à proportion  $\pi$  inconnue)?*

Revoyons à présent les tests couramment en usage, à partir d’un exemple. Supposons que nous ayons recruté  $n_1 = 10$  hommes et  $n_2 = 10$  femmes, tous des sujets échantillonnés au hasard et mutuellement indépendants.



Table 1 ■ Répartition des cas favorables et non-favorables selon l'échantillon

	Cas favorables	Cas non favorables	Totaux
Échantillon 1	$x_1$	$r_1$	$n_1$
Échantillon 2	$x_2$	$r_2$	$n_2$
Totaux	$X$	$R$	$N$

Table 2 ■ Répartition des cas favorables et non-favorables selon l'échantillon

	Couleurs mates	Couleurs brillantes
Hommes	6	4
Femmes	2	8

À chacun, nous demandons si, sur les surfaces de son environnement de travail, il préfère voir des couleurs mates ou des couleurs brillantes. Les réponses (fictives) produisent la compilation du tableau 2.

Nous observons donc  $x_1 = 6, r_1 = 4, x_2 = 2, r_2 = 8, X = x_1 + x_2 = 8, R = r_1 + r_2 = 12, \text{ et } N = 20$ . Au seuil de 5% en mode bilatéral et sur la base de ces données, peut-on croire que l'option de couleurs diffère entre les hommes et les femmes?

**Le test d'association  $\chi^2$ .** Le test du Khi-deux pour un tableau de contingence  $2 \times 2$ , conçu par K. Pearson<sup>1</sup>, se calcule comme suit (Howell, 1998):

$$\chi^2 = \frac{N(x_1 \times r_2 - x_2 \times r_1)^2}{n_1 n_2 X R} \tag{1}$$

Le calcul donne ici:  $\chi^2 = 20 \times (6 \times 8 - 4 \times 2)^2 / (10 \times 10 \times 8 \times 12) \approx 3,33$ . La valeur critique du  $\chi^2$  avec 1 degré de liberté, applicable ici au seuil de 0,05, est égale à 3,84: la valeur observée n'est donc pas significative; la probabilité (implicitement bilatérale) associée est de 0,068.

**Le test d'association  $\chi^2$  avec correction de continuité.** Pour le traitement d'un tableau de contingence  $2 \times 2$ , maints auteurs, dont Siegel (1988) est très représentatif, recommandent l'ajout d'une "correction de continuité", le test ayant la forme modifiée suivante:

$$X_{1/2}^2 = \frac{N(|x_1 \times r_2 - x_2 \times r_1| - 1/2N)^2}{n_1 n_2 X R} \tag{2}$$

Le lecteur vérifiera que la valeur maintenant obtenue égale  $X_c^2 \approx 1,88$ ; cette valeur, non significative en vertu de la même valeur critique, a une probabilité associée de 0,170.

<sup>1</sup>C'est néanmoins R. A. Fisher qui a démontré que la statistique résultante se distribuait comme  $\chi^2$  avec 1 degré de liberté, plutôt qu'avec 3 selon Pearson.

**Le test hypergéométrique de Fisher.** R. A. Fisher, pour traiter les situations statistiques correspondant à un tableau  $2 \times 2$ , a proposé ce qu'on a coutume d'appeler le "test des probabilités exactes de Fisher" (Siegel, 1988), basé sur la distribution hypergéométrique. La probabilité hypergéométrique attachée à un tableau  $2 \times 2$  donné, soit  $x_1, r_1, x_2, r_2$ , est obtenue par:

$$\frac{n_1!n_2!X!R!}{N!x_1!r_1!x_2!r_2!} \tag{3}$$

Pour établir la probabilité unilatérale globale associée à une comparaison donnée, il faut calculer la probabilité associée à la situation observée et lui ajouter les probabilités des situations comparatives à différences plus extrêmes, ce en respectant les totaux marginaux du tableau de départ. Ainsi, dans notre exemple, le tableau observé, 6, 4, 2, 8, donne lieu aux tableaux plus extrêmes 7, 3, 1, 9 et 8, 2, 0, 10, avec probabilités respectives de 0,07502, 0,00953 et 0,00036, lesquelles totalisent  $p = 0,08490$  en mode unilatéral, et  $2 \times p = 0,170$  en mode bilatéral, une valeur comparable à celle dérivée du  $\chi^2$  avec correction, vu auparavant.

**Le test  $t$ .** Sous l'hypothèse nulle, l'estimateur du maximum de vraisemblance de  $\pi$  est la proportion commune de cas favorables,  $\hat{\pi} = X/N$ , avec l'estimateur de variance non biaisé,  $\hat{\pi}(1-\hat{\pi}) \times N/(N-1)$  (Kendall & Stuart, 1977, éq. 33.54), ce qui permet l'approximation par un test  $t$  ayant  $N - 1$  degrés de liberté. La formule suivante implique une correction de continuité qui est justifiée plus loin:

$$t_{1/4} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2} \pm 1/4 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\frac{X R}{N(N-1)} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \tag{4}$$

Pour notre exemple, le calcul donne  $t_{1/4} \approx 1,557$ , avec 19 degrés de liberté. La valeur critique, au seuil bilatéral



de 0,05, est de 2,093. La probabilité bilatérale attachée à ce test est de 0,137.

**Le test z.** Utilisant la proportion commune  $\hat{\pi} = X/N$  et sa variance (asymptotique) associée,  $\hat{\pi}(1 - \hat{\pi})/N$ , l'approximation normale usuelle donne le test  $z$  (Bailargeon & Martin, 1998; McNemar, 1969) suivant:

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{\frac{X_R}{N^2} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad (5)$$

Pour notre cas, la valeur  $z$  calculée égale à peu près 1,826: au carré, cette valeur est de 3,33 et correspond exactement à celle du premier test Khi-deux ci-dessus (1), avec une probabilité bilatérale de 0,068.

**Le test  $z$  avec correction de continuité usuelle.** Le même test admet une correction de continuité "traditionnelle" de  $1/2$  (Guilford, 1965), soit:

$$z_{1/2} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2} \pm 1/2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\frac{X_R}{N^2} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad (6)$$

La valeur  $z_{1/2}$  calculée pour notre exemple est de 1,369, rejoignant au carré la valeur de 1,88 déjà obtenue par le test Khi-deux "corrigé" (2) et sa probabilité bilatérale de 0,170.

**Le test  $z$  avec correction de continuité simple.** L'utilisation d'une correction de continuité a pour but d'améliorer l'équivalence entre les valeurs successives d'une distribution de probabilités discrètes et les aires correspondantes d'une distribution continue utilisée comme modèle d'approximation, ici la loi normale. Or, dans le cas qui nous occupe, la distribution exacte de notre différence  $p_1 - p_2 = x_1/n_1 - x_2/n_2$  a des sauts de continuité soit de  $1/n_1$ , soit de  $1/n_2$ . Si par exemple nous avons  $n_1 = n_2 = n$ , la transition de  $x_1$  à  $x_1 \pm 1$  fera sauter la différence  $x_1/n_1 - x_2/n_2$  d'une valeur de  $\pm 1/n$ , et de même pour chaque transition unitaire de  $x_2$ ; la correction à appliquer à la différence  $x_1/n_1 - x_2/n_2$  serait alors de  $\pm 1/2(1/n)$ , plutôt que de  $\pm 1/2(1/n_1 + 1/n_2)$ . Pour le cas général où  $n_1 \neq n_2$ , il nous faut trouver un intervalle de transition moyen, pour lequel nous proposons  $(1/n_1 + 1/n_2)/2 = 1/n_h$ , où  $n_h$  est la moyenne harmonique de  $n_1$  et  $n_2$ . La correction apparaissant au numérateur prendrait alors la forme indicative:  $p_1 - p_2 \pm 1/2(1/n_h)$  ou, explicitement:  $p_1 - p_2 \pm 1/4(1/n_1 + 1/n_2)$ . L'approximation normale, que nous mettons en lice à l'instar de Pirie et Hamdam (1972; voir aussi Storer and Kim, 1990), sera donc:

$$z_{1/4} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2} \pm 1/4 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\frac{X_R}{N^2} \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad (7)$$

La valeur calculée pour cette approximation sera ici  $z_{1/4} \approx 1,598$ , à quoi se rattache une probabilité bilatérale de 0,110.

**La proposition Kendall-Stuart et Liddell.** Liddell (1978) met en oeuvre une proposition de "test exact" tirée de Kendall and Stuart (1977) et applicable à des échantillons de tailles raisonnables, c'est-à-dire sous-asymptotiques. À leur alinéa 33.22, ces derniers définissent la statistique  $u = x_1/n_1 - x_2/n_2$  et y associent la probabilité binomiale composée:

$$p(x_1, x_2) = b(x_1, n_1 | \pi_1) \times b(x_2, n_2 | \pi_2), \quad (8)$$

la probabilité d'observer une différence (positive) au moins aussi grande que  $u$  étant obtenue par la sommation:

$$p(u' \geq u) = \sum_{t_1, t_2} p(t_1, t_2 | t_1/n_1 - t_2/n_2 \geq u). \quad (9)$$

L'évaluation de l'expression (9) dépend des valeurs réelles de  $\pi_1$  et  $\pi_2$ . Sous l'hypothèse nulle, on doit poser  $\pi_1 = \pi_2 = \pi$ ; cependant, cette valeur unique de  $\pi$  n'est pas connue. Liddell (1978) propose l'estimateur du maximum de vraisemblance basé sur les données, soit la proportion commune  $\hat{\pi} = (n_1 \times p_1 + n_2 \times p_2)/N = X/N$ , et son test pseudo-exact<sup>2</sup> peut alors être effectué, en substituant la valeur  $\hat{\pi}$  pour  $\pi_1$  et  $\pi_2$  dans (8).

Rappelons que, pour notre cas,  $\hat{\pi} = X/N = 0,40$ . De plus, la différence observée  $u$  étant  $6/10 - 2/10 = 0,40$ , les couples  $(t_1, t_2)$  qui engendrent une différence égale ou plus forte sont les 28 suivants: (4:0), (5:1), (5:0), (6:2), (6:1), (6:0), ..., (10:1), (10:0). Pour le premier couple (4:0) et utilisant  $\hat{\pi} = 0,40$ , la probabilité (8) est:

$$p(4, 0) = \binom{10}{4} 0,4^4 0,6^6 \times \binom{10}{0} 0,4^0 0,6^{10} \approx 0,00152.$$

La somme des 28 probabilités associées aux couples énumérés donne 0,05383; en doublant cette valeur pour un test bilatéral, nous obtenons la probabilité globale de 0,108.

<sup>2</sup>L'épithète "exact" est employée ici pour signifier qu'il s'agit d'un calcul de probabilités réelles de l'événement étudié, plutôt que d'un calcul fondé sur une approximation ou un modèle imitatif de l'événement. En statistique inférentielle, ce terme d'"exact" a aussi le sens de "respectant le seuil de signification convenu": p. ex. un test exact au seuil de 5%; nous revenons plus loin sur cette signification, qui possède un certain retentissement dans le cas du test considéré ici.



**Autres tests.** Storer and Kim (1990), qui passent en revue plusieurs formes possibles du test de la différence entre deux proportions, font mention de deux autres procédures. La première, dite “inconditionnelle approximative”, mime la solution de Liddell (1978) citée en (9), donc en utilisant l’estimateur commun  $\hat{\pi}$ , mais en délimitant différemment la zone de rejet. Tandis que, selon Kendall and Stuart (1977) et Liddell (1978), cette zone est constituée par les couples  $(t_1, t_2)$  tels que  $t_1/n_1 - t_2/n_2 \geq x_1/n_1 - x_2/n_2$ , Storer and Kim (1990) l’obtiennent selon  $Q(t_1, t_2) \geq Q(x_1, x_2)$ , où  $Q(t_1, t_2) = (t_1/n_1 - t_2/n_2) / [(t_1 + t_2)(N - t_1 - t_2)N^{-2}(n_1^{-1} + n_2^{-1})]^{1/2}$ . Nous ne cautionnons pas cette proposition,<sup>3</sup> sauf à titre de simple approximation, pour deux raisons: l’ordre imposé par la fonction  $Q$  et la liste de couples  $(t_1, t_2)$  qu’il génère contredisent en partie l’ordre naturel de la différence  $t_1/n_1 - t_2/n_2$ , et le dénominateur de la fonction  $Q$ , qui en principe contient la variance du processus binomial commun aux deux réalisations comparées, varie d’un couple à l’autre (alors qu’il devrait être constant) et contredit l’estimateur  $\hat{\pi}$  employé dans le calcul de probabilité (8).

L’autre procédure évoquée par Storer and Kim (1990) est tirée de Suissa and Shuster (1985) et pompeusement appelée “exacte inconditionnelle”. Elle reprend un calcul similaire à celui de Liddell (d’où l’épithète d’ “exacte”), mais en utilisant comme probabilité de référence l’estimateur  $\pi_{max}$  obtenu comme la valeur de  $\pi$  par laquelle la somme  $\sum_{t_1, t_2} p(t_1, t_2)$  est maximisée et en sélectionnant les couples  $(t_1, t_2)$  selon que  $Q'(t_1, t_2) \geq Q'(x_1, x_2)$ , où  $Q'(t_1, t_2) = (t_1/n_1 - t_2/n_2) / [(t_1(n_1 - t_1)/n_1^3 + t_2(n_2 - t_2)/n_2^3)]^{1/2}$ . La définition de la probabilité de référence nous paraît obscure: la valeur de  $\pi_{max}$  dépend de la liste des couples  $(t_1, t_2)$  retenus dans la sommation, et elle est mal justifiable<sup>4</sup>. Enfin, sauf encore pour un contexte d’approximation, il est difficile d’admettre que non seulement le critère de sélection utilisé (avec  $Q'$ ) diffère du critère évident fourni par  $t_1/n_1 - t_2/n_2 \geq u$  mais en outre qu’il exploite des variances distinctes comme si d’avance on tablait sur la fausseté de l’hypothèse nulle.

Finalement, nous excluons aussi de la compétition l’approximation normale basée sur des variances distinctes, avec ou sans correction de continuité, soit:  $z' = (x_1/n_1 - x_2/n_2) / [x_1 \cdot r_1/n_1^3 + x_2 \cdot r_2/n_2^3]^{1/2}$ : cette statistique, ici, vaut  $z = 2,000$  et est associée à une probabilité bilatérale de 0,0455.

Le lecteur, en considérant les différents tests déjà présentés, aura peut-être entrevu ou imaginé d’autres tests approximatifs, qui seraient ou non des variantes des

tests recensés. Nous ne retiendrons pas ici ces autres tests parce que, en premier lieu, les procédures classiques d’approximation nous semblent suffire et que, en second lieu, nous proposons notre test “exact”, vis-à-vis duquel deux des procédures d’approximation mentionnées font merveille.

**Proposition d’un test exact**

Considérons deux réalisations binomiales, dénotées  $R_1(x_1, n_1)$  et  $R_2(x_2, n_2)$ ; à toutes fins utiles et sans perte de généralité, nous adopterons  $x_1/n_1 \geq x_2/n_2$ . La variable aléatoire  $x_1$  dépend d’un processus binomial  $B(\pi_1, n_1)$  ou, autrement dit,  $x_1 \sim B(\pi_1, n_1)$ ; pour la seconde, on a  $x_2 \sim B(\pi_2, n_2)$ . L’hypothèse nulle allègue que  $\pi_1 = \pi_2 = \pi_0$ , la valeur commune  $\pi_0$  du paramètre  $\pi$  étant inconnue.

Supposons maintenant que nous connaissions  $\pi_0$ . La crédibilité, ou vraisemblance, de l’hypothèse nulle dépend de la facilité avec laquelle les processus binomiaux  $B(\pi_0, n_1)$  et  $B(\pi_0, n_2)$  peuvent produire non seulement les réalisations  $R_1(x_1, n_1)$  et  $R_2(x_2, n_2)$  mais aussi toutes les réalisations parentes dont la différence est égale ou plus grande par rapport à celle observée. À l’instar de Kendall and Stuart (1977), dénotons la différence observée par  $x_1/n_1 - x_2/n_2$ , où  $x_1/n_1 - x_2/n_2 \geq 0$ . Alors, comme en (9) et en admettant  $\pi_0$ , la vraisemblance de l’hypothèse nulle est donnée par:

$$p(\pi_0) = \sum_{t_1, t_2} b(t_1, n_1 | \pi_0) \times b(t_2, n_2 | \pi_0), \quad (10)$$

ce en utilisant toutes les valeurs  $t_1$  et  $t_2$ ,  $0 \leq t_1 \leq n_1$ ,  $t_1, 0 \leq t_2 \leq n_2$ , telles que:

$$t_1/n_1 - t_2/n_2 \geq x_1/n_1 - x_2/n_2. \quad (11)$$

Or, nous ne connaissons pas la valeur de  $\pi_0$  applicable sous l’hypothèse nulle. Nonobstant cette ignorance et le fait que toute valeur de  $\pi$  est possible dans le domaine (0..1), certains intervalles de valeur sont plus plausibles, ou vraisemblables, que d’autres en tenant compte des réalisations  $R_1$  et  $R_2$  enregistrées. En fait, sous l’hypothèse nulle, les réalisations  $R_1$  et  $R_2$  nous informent toutes deux sur  $\pi_0$ . Par exemple, si nous observons  $R_1(10, 10)$ , les intervalles élevés de  $\pi$  ( $\pi \rightarrow 1$ ) sont plus plausibles que les intervalles bas ( $\pi \rightarrow 0$ ). Pour un processus binomial produisant  $t$  cas favorables sur  $n$  occasions, la vraisemblance de chaque valeur de  $\pi$  est chiffrée par la fonction  $V(\pi) = \pi^t(1 - \pi)^{n-t}$ . Supposant les réalisations  $R'_1$  et

<sup>3</sup>Le critère  $Q$  retrouve la liste des couples  $(t_1, t_2)$  énumérés plus haut pour le critère  $t_1/n_1 - t_2/n_2 \geq x_1/n_1 - x_2/n_2$ , sauf le couple (7:3) qu’il refuse et les couples (3:0) et (10:7) qu’il recrute. La probabilité globale qui en résulte est  $p = 0,092$ .

<sup>4</sup>Le critère  $Q'$  appliqué opère ici la même sélection de couples  $(t_1, t_2)$  que le critère  $Q$  (voir note 3). Quant à  $\pi_{max}$ , sa valeur est déterminée non par les proportions observées  $x_1/n_1$  et  $x_2/n_2$  mais par une fonction ( $Q'$ ) de leur différence; son critère, la somme  $\sum_{t_1, t_2} p(t_1, t_2)$ , a une distribution bimodale (pour notre exemple) qui culmine en deux points, soit à  $\pi = 0,2993$  et  $\pi = 0,7007$ ; la somme calculée (après doublement) vaut alors 0,0949.



R'\_2, lesquelles produisent t\_1 + t\_2 cas favorables en n\_1 + n\_2 coups, la vraisemblance de la valeur de pi commune est évidemment donnée par:

V(pi) = pi^{t\_1+t\_2} (1 - pi)^{n\_1+n\_2-t\_1-t\_2} (12)

Cette fonction singe une loi de probabilité bien connue, la loi Bêta (Johnson, Kotz, & Balakrishnan, 1995). Elle culmine sur pi-hat = (t\_1+t\_2)/(n\_1+n\_2), l'estimateur du maximum de vraisemblance, qui est aussi pour ce cas la proportion commune de cas favorables.

Ainsi, toute valeur de pi\_0 étant possible mais non également vraisemblable en considérant les réalisations R\_1 et R\_2, la vraisemblance globale se trouve par la moyenne pondérée de la probabilité (10) sur le domaine de pi, soit 0 < pi < 1, en utilisant la vraisemblance (12) comme fonction de pondération. Ce calcul peut être représenté par:

p\_moy = (int\_0^1 sum\_{t\_1,t\_2} b(t\_1, n\_1 | pi) b(t\_2, n\_2 | pi) V(pi) d pi) / (int\_0^1 V(pi) d pi) (13)

L'expression (13) peut être traitée telle quelle par intégration numérique. On peut aussi la transformer pour en obtenir une expression à calcul exact, par le recours à deux procédés: l'interversion des sommes dans (13) et l'utilisation de la fonction Bêta. Notons d'abord l'égalité formelle: int sum\_{t\_1,t\_2} w(t\_1, t\_2, pi) d pi = sum\_{t\_1,t\_2} int w(t\_1, t\_2, pi) d pi. Deuxièmement, considérant int\_0^1 t^{a-1} (1-t)^{b-1} dt = Bêta(a, b) = (a-1)!(b-1)!/(a+b-1)!, nous obtenons d'abord:

int\_0^1 V(pi) d pi = Bêta(x\_1 + x\_2 + 1, n\_1 + n\_2 - x\_1 - x\_2 + 1)

puis, en appliquant la même équivalence à l'ensemble de l'expression (13) et en utilisant N = n\_1 + n\_2, X = x\_1 + x\_2, nous obtenons enfin:

p\_moy = ((N+1) \* binom(N, X)) / (2N+1) \* sum\_{t\_1,t\_2} (binom(n\_1, t\_1) \* binom(n\_2, t\_2)) / binom(X+t\_1+t\_2, t\_1+t\_2) (14)

expression qu'on peut redonner en fonctions factorielles, soit:

p\_moy = ((N+1)! \* n\_1! \* n\_2!) / ((2N+1)! \* X! \* (N-X)!) \* sum\_{t\_1,t\_2} ((X+t\_1+t\_2)! \* (2N-X-t\_1-t\_2)!) / (t\_1! \* (n\_1-t\_1)! \* t\_1! \* (n\_2-t\_2)!) (15)

Cette expression fournit, d'après nous, une évaluation exacte de la vraisemblance de l'hypothèse nulle à partir de la différence observée entre les proportions x\_1/n\_1 et x\_2/n\_2.

L'évaluation de (14) ou (15), sujette à (11), à partir des données de notre exemple, fournit la valeur p\_moy approx 0,05050, ce qui, en mode bilatéral, revient à 0,101.

Répertoire des tests à comparer

Non seulement y a-t-il de nombreux tests en lice pour aider à la décision sur la différence entre deux proportions indépendantes, mais ces tests reposent sur des bases variées et ils produisent des décisions souvent disparates.

Notre test exact, correspondant à l'expression (14), a pour mérite d'exploiter le principe d'analyse proposé par Kendall et Stuart (1977), à l'instar de Liddell (1978); de ne faire intervenir aucune approximation, ni comme modèle de probabilité ni pour l'estimation de la probabilité de référence pi\_0, et enfin de fournir une vraisemblance de l'hypothèse d'égalité: pi\_1 = pi\_2, basée réalistement sur les proportions observées x\_1/n\_1 et x\_2/n\_2. Nous retenons ici ce test comme étalon dans l'examen comparatif des autres tests.

Le test hypergéométrique de Fisher, souvent appliqué pour la situation présente, suppose néanmoins que les totaux marginaux sont fixes (ou fixés d'avance); or, si dans la comparaison de deux groupes les tailles n\_1 et n\_2 sont connues et ne dépendent pas de l'expérience en jeu, leurs réponses x\_1 et x\_2 sont des variables aléatoires, de même que les totaux X et R qui en découlent. Il s'ensuit que le test de Fisher n'est pas approprié ici, sauf comme approximation et sur un mode asymptotique (Kendall & Stuart, 1977).

L'approximation par la distribution t de Student (4), exploitant la proportion commune X/N = pi-hat comme estimateur de pi\_0 et l'estimateur de variance non biaisé, pi-hat(1 - pi-hat)N/(N - 1), n'est pas répertoriée dans la documentation trouvée. Nous l'utilisons ici avec la correction de continuité "simple" au numérateur, +/-1/4(1/n\_1 + 1/n\_2).

L'approximation normale de base (5), symbolisée "z", correspond algébriquement au test du Khi-deux sans correction de continuité (1), une version recommandée par exemple par Conover (1974). Nous retenons ici la forme "z".

Le test par approximation normale comporte le plus souvent une correction de continuité, soit la correction +/-1/2(1/n\_1 + 1/n\_2), qui donne la forme "z\_{1/2}", algébriquement équivalente au chi\_c^2 (2), ou la correction +/-1/4(1/n\_1 + 1/n\_2), fournissant la forme "z\_{1/4}", qu'on retrouve dans un format chi^2 chez Pirie and Hamdam (1972) puis Storer and Kim (1990). Encore ici, nous retenons les formes "z\_{1/2}" et "z\_{1/4}".

La sixième procédure retenue est celle de Liddell (1978), étudiée aussi par Storer and Kim (1990), et qui est une réduction de notre test exact, la réduction consistant à ne retenir de la fonction de vraisemblance (12) que son



maximum et à appliquer  $\pi_0 = \hat{\pi}$  plutôt que le domaine entier  $\pi_0 \in (0..1)$ ; il s'agit donc, quant à nous, d'un test approximatif.

### Comparaison des tests et conclusion

Le tableau 3 rapporte les résultats d'une comparaison des sept procédures de test retenues, les tests 1 à 6 étant confrontés à notre test binomial dit "exact". Prenons par exemple la ligne  $n_1 = n_2 = 5$ , qui donne lieu à  $1/2 \times 6 \times 5 = 15$  différences de type  $x_1/n_1 - x_2/n_2 > 0$ . Pour ce contexte et au seuil unilatéral de 0,05, le test 3 s'accorde 15 fois sur 15 avec le test exact (0 discordance), alors que les tests 5 et 6 ratent chacun une décision positive (pour  $x_1 : x_2 = 4 : 1$ ). Quant aux tests 1, 2 et 4, leurs trois discordances sont communes et consistent à rester insensibles à des différences qui seraient significatives (pour  $x_1 : x_2 = 3:0, 4:1$  et  $2:5$ ).

Les conclusions suivantes s'imposent:

1. Les tests 5 (approximation  $z_{1/4}$ ) et 6 (Liddell) concordent presque parfaitement avec le test exact et entre eux. Quand discordance il y a, elle se manifeste par un léger conservatisme pour le test 5 versus une légère libéralité pour le test 6. Prenons des échantillons de tailles  $n_1 = 20$  et  $n_2 = 40$ , avec  $1/2 \times 21 \times 40 = 420$  différences positives: pour ce contexte et au seuil de 0,05, nous notons seulement 4 cas de discordance d'un test ou l'autre, soit 0,9% des cas examinés.
2. Les tests 1 (probabilités hypergéométriques de Fisher) et 4 ( $z_{1/2}$ ) pèchent tous deux par conservatisme et, de fait, semblent covarier presque parfaitement. Un examen des cas de discordance réciproque montre en effet que le test  $z_{1/2}$  approxime très efficacement le test de Fisher.
3. Le test 3 ( $z$ ), trop libéral, contraste avec le test 2 ( $t_{1/4}$ ), trop conservateur, les deux étant comme assis entre deux chaises.
4. Les discordances tendent à augmenter en fréquence à mesure que les tailles  $n_1$  et  $n_2$  croissent, cette tendance affectant surtout les tests 1 à 4 et pratiquement pas les tests 5 et 6.

L'étroite correspondance constatée entre le test hypergéométrique de Fisher (test 1) et le  $z_{1/2}$  ou  $\chi_c^2$  (test 4) illustre, d'après nous, que ce dernier ne convient pas à la comparaison de deux proportions indépendantes, le principe du test de Fisher imposant un taux de succès global prédéterminé.

L'effet de turbulence constaté plus haut (en conclusion 4), et affectant les tests 1 à 4, pourrait être interprété a contrario, comme suit. En admettant que le test 4 (ou 3) soit juste et notre test "exact" seulement approximatif, ce serait ce dernier, avec ses compagnons 5 et 6, qui mani-

festeraient de la turbulence comparativement au test indiqué. Cependant ce point de vue nous semble difficilement tenable, du fait que le test "exact" proposé ne comporte rien d'approximatif dans son calcul ou dans ses paramètres, et qu'il est alors peu plausible que son résultat affiche une dérive lorsque les tailles  $n_j$  augmentent.

En conclusion sommaire, donc, nous proposons que notre test "exact" (14) soit la référence pour la comparaison de deux proportions indépendantes, nos données d'étude montrant que l'approximation " $z_{1/4}$ " ou son équivalent Khi-deux lui sont un excellent substitut.

### Note de l'auteur

Une version antérieure de cet article a paru en 2005 dans *Lettres Statistiques*, vol. 12, p. 1-13.

### References

- Baillargeon, G. & Martin, L. (1998). *Méthodes quantitatives et analyse de données en sciences humaines*. Trois-Rivières: SMG.
- Conover, W. J. (1974). Some reasons for not using the yates continuity correction on  $2 \times 2$  contingency tables. *Journal of the American Psychological Association*, 69, 374-382.
- Guilford, J. P. (1965). *Fundamental statistics in psychology and education (4e édition)*. New York: McGraw-Hill.
- Howell, D. C. (1998). *Méthodes statistiques en sciences humaines*. Paris: De Boeck Université.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions (vol. 2) (2e édition)*. New York: Wiley.
- Kendall, M. G. & Stuart, A. (1977). *The advanced theory of statistics. Vol. 1*.
- Liddell, D. (1978). Practical tests of  $2 \times 2$  contingency tables. *The Statistician*, 25, 295-304.
- McNemar, Q. (1969). *Psychological statistics (4e édition)*. New York: Wiley.
- Pirie, W. R. & Hamdam, M. A. (1972). Some revised continuity corrections for discrete distributions. *Biometrics*, 28, 693-701.
- Siegel, S. (1988). *Nonparametric statistics for the behavioral sciences (2e édition)* New York: McGraw-Hill.
- Storer, B. E. & Kim, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, 85, 146-156.
- Suissa, S. & Shuster, J. J. (1985). Exact unconditional sample sizes for the  $2 \times 2$  binomial trial. *Journal of the Royal Statistical Society (Series A)*, 148, 317-327.



**Table 3** ■ Comparaison des décisions rendues par six tests de la différence entre deux proportions par rapport au test exact (éq. 14)

$n_1$	$n_2$	Seuil unilatéral $\alpha = 0,05$						Seuil unilatéral $\alpha = 0,01$					
		Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
5	5	3/0	3/0	0/0	3/0	1/0	1/0	2/0	3/0	0/0	2/0	2/0	0/0
10	10	2/0	2/0	0/5	2/0	0/0	0/0	1/0	3/0	0/4	3/0	0/0	0/2
15	15	7/0	5/0	0/6	9/0	1/0	0/0	5/0	5/0	0/4	7/0	0/0	0/0
20	20	9/0	2/0	0/8	9/0	0/0	0/0	5/0	5/0	0/10	9/0	0/2	0/2
25	25	6/0	2/0	0/16	8/0	0/2	0/2	9/0	7/0	0/10	11/0	0/2	0/2
30	30	18/0	4/0	0/12	18/0	0/0	0/0	6/0	2/0	0/20	7/0	0/4	0/4
5	10	4/0	4/0	0/1	4/0	1/0	0/0	3/0	4/0	0/0	3/0	1/0	0/0
5	20	7/0	3/0	0/2	7/0	3/0	0/1	5/0	3/0	0/2	5/0	2/1	0/2
5	30	11/0	5/0	0/2	12/0	5/0	0/0	10/0	6/0	1/3	8/0	4/0	0/1
10	20	7/0	6/0	0/5	7/0	2/1	0/2	5/0	5/0	0/6	5/0	2/0	0/0
10	30	12/0	7/0	0/6	10/0	4/0	0/0	9/0	7/0	0/6	10/0	3/0	1/1
20	30	19/0	10/0	0/6	18/0	7/0	0/1	14/0	13/0	0/5	15/0	3/0	0/1
20	40	15/0	4/0	0/15	15/0	3/1	0/4	3/0	8/0	0/9	15/0	4/0	1/2

*Note.* Note: La notation “a / b” indique le nombre de décisions discordantes rendues par le test indiqué comparative-ment à notre test exact: “a” désigne les fois où le test indiqué n’est pas significatif (au seuil prescrit) et le nôtre l’est, tandis que “b” désigne les fois où le test indiqué est significatif et le nôtre non. Le nombre d’occasions de discordance (qui peut servir de dénominateur aux nombres présentés) est égal à  $\frac{1}{2}[\min(n_1, n_2) + 1] \cdot \max(n_1, n_2)$ .

Les tests sont: Test 1 - probabilités hypergéométriques de Fisher; Test 2 - approximation  $t_{1/4}$ ; Test 3 - approximation  $z$ ; Test 4 - approximation  $z_{1/2}$ ; Test 5 - approximation  $z_{1/4}$ ; Test 6 - probabilités binomiales de Liddell.

**Citation**

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(2), 141–147. doi:10.20982/tqmp.13.2.p141

Copyright © 2017, 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: 02/05/2017 ~ Accepted: 12/05/2017