

Test de pourcentage de défectueux à 2 échantillons

Généralités

Un test à 2 proportions permet de déterminer si deux proportions sont significativement différentes. Dans le domaine de l'analyse de la qualité, ce test est souvent utilisé lorsqu'un produit ou un service est caractérisé par le statut défectueux ou non défectueux, afin de déterminer si le pourcentage de défectueux diffère de façon significative dans des échantillons collectés à partir de deux procédés indépendants.

L'Assistant de Minitab est doté d'un test de pourcentage de défectueux à 2 échantillons. Les données collectées pour le test représentent le nombre de défectueux dans chacun des deux échantillons indépendants, considéré comme la valeur observée d'une variable aléatoire obéissant à une loi binomiale. L'Assistant utilise des méthodes exactes pour calculer les résultats du test d'hypothèse ; aussi, le taux d'erreur de 1^{ère} espèce réel est a priori proche du seuil de signification (α) spécifié pour le test et aucune étude plus approfondie n'est requise. En revanche, l'Assistant utilise une méthode d'approximation selon la loi normale pour calculer l'intervalle de confiance (IC) pour la différence entre les pourcentages de défectueux et une fonction puissance théorique du test d'approximation selon la loi normale pour effectuer l'analyse de puissance et d'effectif d'échantillon. Etant donné qu'il s'agit de méthodes par approximation, il est nécessaire d'en évaluer l'exactitude.

Dans cette étude, nous évaluons les conditions dans lesquelles les intervalles de confiance obtenus par approximation sont exacts. Nous étudions également la méthode utilisée pour évaluer la puissance et l'effectif d'échantillon pour le test de pourcentage de défectueux à 2 échantillons, et comparons la puissance théorique obtenue avec la méthode par approximation à la puissance réelle du test exact. Enfin, nous décrivons les vérifications des données suivantes, qui sont automatiquement effectuées et affichées dans le rapport de l'Assistant, et nous expliquons comment elles influent sur les résultats de l'analyse :

- Validité de l'IC
- Effectif d'échantillon

Le test de pourcentage de défectueux à 2 échantillons dépend également d'autres hypothèses. Pour plus de détails, reportez-vous à l'Annexe A.

Méthodes de test de pourcentage de défectueux à 2 échantillons

Exactitude des intervalles de confiance

Bien que l'Assistant utilise le test exact de Fisher pour déterminer si le pourcentage de défectueux des deux échantillons diffère de manière significative, l'intervalle de confiance pour la différence s'appuie sur la méthode d'approximation selon la loi normale. D'après la règle générale énoncée dans la plupart des manuels de statistiques, cet intervalle de confiance par approximation est exact si le nombre de défectueux et de non-défectueux observés dans chaque échantillon est d'au moins 5.

Objectif

Nous souhaitons déterminer les conditions dans lesquelles les intervalles de confiance fondés sur l'approximation selon la loi normale sont exacts. De manière plus spécifique, nous souhaitons évaluer la façon dont la règle générale relative au nombre de défectueux et de non-défectueux dans chaque échantillon influe sur l'exactitude des intervalles de confiance par approximation.

Méthode

La formule utilisée pour calculer l'intervalle de confiance pour la différence entre les deux proportions et la règle générale qui en garantit l'exactitude sont décrites dans l'Annexe D. Nous décrivons également une règle modifiée et moins rigoureuse, que nous avons développée au cours de notre étude.

Nous avons effectué des simulations pour évaluer l'exactitude de l'intervalle de confiance obtenu par approximation sous différentes conditions. Pour ce faire, nous avons généré des paires d'échantillons aléatoires d'effectifs variés à partir de plusieurs populations de Bernoulli. Pour chaque type de population de Bernoulli, nous avons calculé un intervalle de confiance par approximation pour la différence entre les deux proportions de chaque paire de 10 000 répliques d'échantillon de Bernoulli. Nous avons ensuite calculé la proportion des 10 000 intervalles qui contenaient la différence réelle entre les deux proportions, aussi appelée probabilité de couverture simulée. Si l'intervalle par approximation est exact, la probabilité de couverture simulée sera proche de la probabilité de couverture cible de 0,95. Pour évaluer l'exactitude de l'intervalle par approximation pour la règle initiale et la règle modifiée sur le nombre minimal de défectueux et de non-défectueux nécessaires dans chaque échantillon, nous avons aussi calculé le pourcentage des 10 000 paires d'échantillons pour lesquels chaque règle était satisfaite. Pour plus de détails, reportez-vous à l'Annexe D.

Les résultats

L'intervalle de confiance obtenu par approximation pour la différence entre deux proportions est généralement exact lorsque les échantillons sont suffisamment grands, c'est-à-dire lorsque le nombre observé de défectueux et de non-défectueux dans chaque échantillon est

d'au moins 5. Nous avons donc adopté cette règle pour tester la validité de l'IC dans le rapport. Bien que cette règle offre généralement de bons résultats, elle peut se montrer trop prudente dans certains cas ou un peu trop souple lorsque les deux proportions sont proches de 0 ou de 1. Pour plus de détails, reportez-vous à l'Annexe D et à la section Vérification des données.

Efficacité de la fonction puissance théorique

L'Assistant effectue le test d'hypothèse pour comparer deux proportions de population de Bernoulli (pourcentages de défectueux dans deux échantillons) à l'aide du test de Fisher. Toutefois, étant donné que la fonction puissance de ce test exact est difficile à dériver, il nous faut l'obtenir par une approximation, à l'aide de la fonction puissance théorique du test correspondant d'approximation selon la loi normale.

Objectif

Nous souhaitons déterminer si la fonction puissance théorique issue du test d'approximation selon la loi normale était adaptée pour évaluer la puissance et l'effectif d'échantillon nécessaires pour le test de pourcentage de défectueux à 2 échantillons dans l'Assistant. Pour ce faire, il nous fallait déterminer si cette fonction puissance théorique reflétait la puissance réelle du test exact de Fisher.

Méthode

La méthodologie utilisée pour le test exact de Fisher, notamment pour le calcul de sa valeur de p , est décrite en détail dans l'Annexe B. La fonction puissance théorique issue du test d'approximation selon la loi normale est définie dans l'Annexe C. Partant de ces définitions, nous avons effectué des simulations pour estimer les niveaux de puissance réelle (aussi appelés niveaux de puissance simulée) du test exact de Fisher lorsqu'il est utilisé pour analyser la différence entre les pourcentages de défectueux de deux échantillons.

Pour ce faire, nous avons généré des paires d'échantillons aléatoires d'effectifs variés à partir de plusieurs populations de Bernoulli. Pour chaque catégorie de population de Bernoulli, nous avons effectué le test exact de Fisher sur chaque paire de 10 000 répliques d'échantillon. Pour chaque effectif d'échantillon, nous avons calculé la puissance simulée du test pour la détection d'une différence donnée, exprimée comme la fraction des 10 000 paires d'échantillons pour laquelle le test est significatif. Pour effectuer la comparaison, nous avons également calculé la puissance théorique correspondante pour le test d'approximation selon la loi normale. Si l'approximation est bonne, les niveaux de puissance théorique et simulée doivent être proches. Pour plus de détails, reportez-vous à l'Annexe E.

Les résultats

Nos simulations ont montré que, de manière générale, la fonction puissance théorique du test d'approximation selon la loi normale et la fonction puissance simulée du test exact de Fisher sont approximativement égales. Par conséquent, l'Assistant utilise la fonction puissance théorique du test d'approximation selon la loi normale pour calculer les effectifs

d'échantillons nécessaires pour détecter des différences pratiques importantes dans le cadre du test exact de Fisher.

Vérification des données

Validité de l'IC

Etant donné que le test de pourcentage de défectueux à 2 échantillons utilise un test exact afin d'évaluer la différence entre les pourcentages de défectueux, le nombre de défectueux et de non-défectueux présents dans chaque échantillon n'a pas beaucoup d'effet sur son exactitude. Toutefois, l'intervalle de confiance pour la différence entre les pourcentages de défectueux repose sur une approximation selon la loi normale. Lorsque le nombre de défectueux et de non-défectueux présents dans chaque échantillon augmente, l'exactitude de l'intervalle de confiance obtenu par approximation augmente également (voir l'Annexe D).

Objectif



Nous souhaitons déterminer si le nombre de défectueux et de non-défectueux présents dans les échantillons était suffisant pour garantir l'exactitude de l'intervalle de confiance obtenu par approximation pour la différence entre les pourcentages de défectueux.

Méthode

Nous avons utilisé la règle générale énoncée dans la plupart des manuels de statistiques. Lorsque chaque échantillon contient au moins 5 défectueux et 5 non-défectueux, l'intervalle de confiance obtenu par approximation pour le test de pourcentage de défectueux à 2 échantillons est exact. Pour plus de détails, reportez-vous à la section Méthodes de test de pourcentage de défectueux à 2 échantillons ci-dessus.

Les résultats

Comme le montrent les simulations récapitulées dans la section Méthodes de test de pourcentage de défectueux à 2 échantillons, l'exactitude de l'intervalle de confiance dépend du nombre minimal de défectueux et de non-défectueux présents dans chaque échantillon. Par conséquent, l'Assistant affiche les indicateurs d'état suivants dans le rapport pour évaluer l'exactitude de l'intervalle de confiance pour la différence entre deux pourcentages de défectueux.

Etat	Condition
	Les deux échantillons présentent au moins 5 défectueux et 5 non-défectueux. L'intervalle de confiance pour la différence devrait être exact.
	Le nombre de défectueux ou de non-défectueux est inférieur à 5 dans au moins un échantillon. L'intervalle de confiance pour la différence risque de ne pas être exact.

Effectif d'échantillon

Par définition, un test d'hypothèse statistique vise à collecter des preuves permettant de rejeter l'hypothèse nulle de "non-différence". Lorsque l'échantillon est trop faible, la puissance du test peut ne pas être adaptée pour détecter une différence existante, ce qui entraîne une erreur de 2^{ème} espèce. Il est donc essentiel de s'assurer que les effectifs d'échantillons sont suffisamment grands pour détecter des différences importantes dans la pratique avec une probabilité élevée.

Objectif

Si les données ne permettent pas de rejeter l'hypothèse nulle, il nous faut déterminer si les effectifs d'échantillons sont suffisamment grands pour que le test détecte des différences pratiques avec une probabilité élevée. Même si la planification des effectifs d'échantillons vise à garantir que les effectifs d'échantillons sont suffisamment grands pour détecter d'importantes différences avec une probabilité élevée, ces effectifs ne doivent pas être grands au point que des différences sans importance deviennent statistiquement significatives avec une probabilité élevée.






Méthode

L'analyse de puissance et d'effectif d'échantillon pour le test de pourcentage de défectueux à 2 échantillons s'appuie sur la fonction puissance théorique du test d'approximation selon la loi normale. Cette fonction puissance fournit une bonne estimation de la puissance réelle du test exact de Fisher (reportez-vous aux résultats de simulation récapitulés dans la section Méthodes de test de pourcentage de défectueux à 2 échantillons, dans la partie relative à l'efficacité de la fonction puissance théorique). La fonction puissance théorique peut être exprimée comme fonction de la différence cible dans le pourcentage de défectueux et du pourcentage global de défectueux présents dans les échantillons combinés.

Les résultats

Lorsque les données ne fournissent pas suffisamment de preuves invalidant l'hypothèse nulle, l'Assistant utilise la fonction puissance du test d'approximation selon la loi normale pour calculer les différences pratiques pouvant être détectées avec une probabilité de 80 % et de 90 % pour l'effectif d'échantillon donné. De plus, si l'utilisateur indique une différence pratique spécifique présentant un intérêt particulier, l'Assistant utilise la fonction puissance du test d'approximation selon la loi normale pour calculer les effectifs d'échantillons qui offrent une probabilité de 80 % et de 90 % de détecter cette différence.

Pour faciliter l'interprétation des résultats, le rapport de l'Assistant pour le test du pourcentage de défectueux à 2 échantillon affiche les indicateurs d'état suivants lors du test de la puissance et de l'effectif d'échantillon :

Etat	Condition
	<p>Le test détecte une différence entre les pourcentages de défectueux, la puissance n'est donc pas un problème.</p> <p>OU</p> <p>La puissance est suffisante. Le test n'a pas détecté de différence entre les pourcentages de défectueux, mais l'échantillon est suffisamment grand pour fournir une probabilité d'au moins 90 % de détecter la différence donnée (puissance ≥ 90).</p>
	<p>La puissance est peut-être suffisante. Le test n'a pas détecté de différence entre les pourcentages de défectueux, mais l'échantillon est suffisamment grand pour fournir une probabilité de 80 % à 90 % de détecter la différence donnée ($0,80 \leq$ puissance $< 0,90$). L'effectif d'échantillon nécessaire pour atteindre une puissance de 90 % est indiqué.</p>
	<p>Il se peut que la puissance ne soit pas suffisante. Le test n'a pas détecté de différence entre les pourcentages de défectueux, et l'échantillon est suffisamment grand pour fournir une probabilité de 60 % à 80 % de détecter la différence donnée ($0,60 \leq$ puissance $< 0,80$). Les effectifs d'échantillon nécessaires pour atteindre une puissance de 80 % et de 90 % sont indiqués.</p>
	<p>La puissance n'est pas suffisante. Le test n'a pas détecté de différence entre les pourcentages de défectueux, et l'échantillon n'est pas suffisamment grand pour fournir une probabilité d'au moins 60 % de détecter la différence donnée (puissance $< 0,60$). Les effectifs d'échantillon nécessaires pour atteindre une puissance de 80 % et de 90 % sont indiqués.</p>
	<p>Le test n'a pas détecté de différence entre les pourcentages de défectueux. Vous n'avez pas indiqué de différence pratique à détecter. Selon vos données, le rapport peut indiquer les différences que vous pourriez détecter avec une probabilité de 80 % et de 90 %, en fonction de l'effectif de votre échantillon et de la valeur d'alpha.</p>

Références

Arnold, S.F. (1990), *Mathematical statistics*, Englewood Cliffs, NJ : Prentice Hall, Inc.

Casella, G. et Berger, R.L. (1990), *Statistical inference*, Pacific Grove, CA : Wadsworth, Inc.

Annexe A : hypothèses supplémentaires pour le test du pourcentage de défectueux à 2 échantillons

Le test de pourcentage de défectueux à 2 échantillons repose sur les hypothèses suivantes :

- Les données de chaque échantillon sont constituées de n éléments distincts, chacun d'eux étant classé comme défectueux ou non défectueux.
- La probabilité qu'un élément soit défectueux est la même pour chaque élément au sein d'un échantillon.
- La probabilité qu'un élément soit défectueux n'est pas influencée par le fait que l'élément précédent soit défectueux ou non.

Ces hypothèses ne peuvent pas être vérifiées dans les vérifications des données du rapport, car les données entrées pour ce test sont des données récapitulatives et non des données brutes.

Annexe B : test exact de Fisher

Supposons que nous observons deux échantillons aléatoires indépendants X_1, \dots, X_{n_1} et Y_1, \dots, Y_{n_2} obéissant à des lois de Bernoulli, tel que :

$$p_1 = \Pr(X_i = 1) = 1 - \Pr(X_i = 0) \text{ et } p_2 = \Pr(Y_j = 1) = 1 - \Pr(Y_j = 0)$$

Dans les sections suivantes, nous décrivons les procédures pour la création d'inférences sur la différence entre les proportions $\delta = p_1 - p_2$.

Formule D1 : test exact de Fisher et valeur de p

Arnold (1994) offre une description du test exact de Fisher. Nous décrivons brièvement ce test dans notre étude.

Soit V le nombre de succès obtenus dans le premier échantillon et $v = n_1 \hat{p}_1$ le nombre observé de succès dans le premier échantillon lors de la réalisation d'une expérience. En outre, soit W le nombre total de succès obtenus dans les deux échantillons et $w = n_1 \hat{p}_1 + n_2 \hat{p}_2$ les succès observés lors de la réalisation d'une expérience. Notez que \hat{p}_1 et \hat{p}_2 sont les estimations ponctuelles d'échantillon de p_1 et p_2 .

Sous l'hypothèse nulle selon laquelle $\delta = p_1 - p_2 = 0$, la loi de distribution conditionnelle de V sachant W est la loi hypergéométrique avec la fonction de masse suivante :

$$f(v|w) = \frac{\binom{n_1}{v} \binom{n_2}{w-v}}{\binom{n_1 + n_2}{w}}$$

Soit $F(v|w)$ la fonction de distribution cumulative de la loi. Les valeurs de p des tests unilatéraux et bilatéral sont les suivantes :

- **Avec l'hypothèse alternative $\delta < 0$ (soit $p_1 < p_2$)**

La valeur de p est calculée comme suit : $F(v|w)$, où v est la valeur observée de V ou le nombre observé de succès dans le premier échantillon, et w la valeur observée de W ou le nombre observé de succès dans les deux échantillons.

- **Avec l'hypothèse alternative $\delta > 0$ (soit $p_1 > p_2$)**

La valeur de p est calculée comme suit : $1 - F(v - 1|w)$, où v est la valeur observée de V ou le nombre observé de succès du premier échantillon et w , la valeur observée de W ou le nombre observé de succès des deux échantillons.

- Avec l'hypothèse alternative $\delta \neq 0$ (soit $p_1 \neq p_2$)

La valeur de p est calculée avec l'algorithme suivant, où m est le mode de loi hypergéométrique décrit ci-dessus.

- Si $v < m$, la valeur de p est calculée comme suit : $1 - F(y - 1|w) + F(v|w)$, où v et w sont tels que définis précédemment et $y = \min\{k \geq m: f(k|w) \leq f(v|W)\}$
- Si $v = m$, la valeur de p est de 1,0
- Si $v > m$, la valeur de p est calculée comme suit : $1 - F(v - 1|w) + F(y|w)$, où v et w sont tels que définis précédemment et $y = \max\{k \leq m: f(k|w) \leq f(v|W)\}$

Annexe C : fonction puissance théorique

Pour comparer deux proportions (ou, plus précisément, deux pourcentages de défectueux), nous utilisons le test exact de Fisher, comme nous l'avons décrit dans l'Annexe B. Le calcul d'une fonction puissance théorique pour ce test étant trop complexe, nous utilisons une fonction puissance par approximation. Plus spécifiquement, nous utilisons la fonction puissance du test bien connu d'approximation selon la loi normale pour deux proportions afin d'obtenir une approximation de la puissance du test exact de Fisher.

La fonction puissance du test d'approximation selon la loi normale pour le test bilatéral est la suivante :

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{es}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{es}\right)$$

où $\delta = p_1 - p_2$,

$$es = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

et $p_c = (n_1p_1 + n_2p_2)/(n_1 + n_2)$.

Pour le test de l'hypothèse nulle $p_1 = p_2$ contre l'hypothèse alternative $p_1 > p_2$, la fonction puissance est la suivante :

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{es}\right)$$

Pour le test de l'hypothèse nulle $p_1 = p_2$ contre l'hypothèse alternative $p_1 < p_2$, la fonction puissance est la suivante :

$$\pi(n_1, n_2, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{es}\right)$$

Annexe D : intervalles de confiance par approximation

Formule D1 : calcul d'un intervalle de confiance par approximation pour la différence entre deux proportions

Un intervalle de confiance asymptotique à $100(1 - \alpha)\%$ pour $\delta = p_1 - p_2$ issu d'une approximation selon la loi normale est exprimé comme suit :

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$

Il existe une règle générale connue permettant d'évaluer la fiabilité de cet intervalle de confiance par approximation : $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$. En d'autres termes, l'intervalle de confiance est exact si le nombre observé de succès et de défaillances dans chaque échantillon est d'au moins 5.

Remarque : dans cette section et les suivantes, nous exprimons la règle de l'intervalle de confiance sous sa forme la plus courante, en fonction du nombre de succès et de défaillances dans chaque échantillon. Un succès est un événement d'intérêt et une défaillance le complément de l'événement d'intérêt. Par conséquent, dans le contexte spécifique du test de pourcentage de défectueux à 2 échantillons, le nombre de "succès" équivaut au nombre de défectueux et le nombre de "défaillances" à celui de non-défectueux.

Formule D2 : règles pour les intervalles de confiance par approximation

La règle générale utilisée pour les intervalles de confiance fondés sur l'approximation selon la loi normale veut que les intervalles de confiance soient exacts si $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$. En d'autres termes, le niveau de confiance réel de l'intervalle est égal ou approximativement égal au niveau de confiance cible si chaque échantillon contient au moins 5 succès (défectueux) et 5 défaillances (non-défectueux).

La règle est exprimée en fonction des proportions estimées de succès et de défaillances, par opposition aux proportions réelles, inconnues dans la pratique. Toutefois, dans les paramètres théoriques où les proportions réelles sont supposées ou connues, la règle peut être exprimée directement à partir des proportions réelles. Dans ces cas-ci, nous pouvons directement déterminer dans quelle mesure le nombre de succès et de défaillances prévu, n_1p_1 , n_2p_2 , $n_1(1 - p_1)$ et $n_2(1 - p_2)$, influence sur la probabilité de couverture réelle de l'intervalle de confiance pour la différence entre les proportions.

Nous pouvons évaluer la probabilité de couverture réelle en testant un grand nombre de paires d'échantillons d'effectifs n_1 et n_2 issus des deux populations de Bernoulli avec des probabilités de succès p_1 et p_2 . La probabilité de couverture réelle est ensuite calculée

comme la fréquence relative à laquelle les paires d'échantillons génèrent des intervalles de confiance qui contiennent la différence réelle entre les deux proportions. Si la probabilité de couverture réelle est exacte lorsque $n_1 p_1 \geq 5$, $n_2 p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ et $n_2(1 - p_2) \geq 5$, alors, d'après la loi des grands nombres, la probabilité de couverture est exacte lorsque $n_1 \hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2 \hat{p}_2 \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$. Par conséquent, lorsque les niveaux de confiance réel et cible sont proches, on pourrait s'attendre à ce qu'une grande proportion des paires d'échantillons générées à partir des deux populations de Bernoulli obéissent aux conditions $n_1 \hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2 \hat{p}_2 \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$ lorsque la règle est valide. Dans la simulation suivante, nous désignons cette règle comme la règle 1.

En outre, au cours de cette étude, nous avons remarqué dans de nombreux cas que si $n_1 p_1 \geq 5$ et $n_2 p_2 \geq 5$ ou si $n_1(1 - p_1) \geq 5$ et $n_2(1 - p_2) \geq 5$, la probabilité de couverture simulée de l'intervalle est proche de la couverture cible. Nous en avons tiré une autre règle, plus souple, selon laquelle les intervalles de confiance obtenus par approximation sont exacts lorsque $n_1 \hat{p}_1 \geq 5$ et $n_2 \hat{p}_2 \geq 5$ ou lorsque $n_1(1 - \hat{p}_1) \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$. Dans la simulation suivante, nous appelons cette règle modifiée règle 2.

Simulation D1 : évaluation de l'exactitude des intervalles de confiance par approximation

Nous avons effectué des simulations pour évaluer les conditions dans lesquelles l'intervalle de confiance par approximation pour la différence entre deux proportions est exact. Plus précisément, nous avons étudié l'exactitude de l'intervalle par rapport aux règles générales suivantes :

Règle 1 (initiale) $n_1 p_1 \geq 5$, $n_2 p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ et $n_2(1 - p_2) \geq 5$

Règle 2 (modifiée) $n_1 \hat{p}_1 \geq 5$ et $n_2 \hat{p}_2 \geq 5$ OU $n_1(1 - \hat{p}_1) \geq 5$ et $n_2(1 - \hat{p}_2) \geq 5$

Dans chaque expérience, nous avons généré 10 000 paires d'échantillons à partir de paires de populations de Bernoulli définies par les proportions suivantes :

- **Proportions A : p_1 et p_2 sont proches de 1,0 (ou de 0).** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,8$ et $p_2 = 0,9$.
- **Proportions B : p_1 et p_2 sont proches de 0,5.** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,4$ et $p_2 = 0,55$.
- **Proportions C : p_1 est proche de 0,5 et p_2 est proche de 1,0.** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,4$ et $p_2 = 0,9$.

La classification des proportions décrite ci-dessus repose sur l'approximation selon la loi normale de DeMoivre-Laplace pour la loi binomiale à partir de laquelle les intervalles de confiance par approximation sont dérivés. Cette approximation selon la loi normale est exacte lorsque l'effectif d'échantillon de Bernoulli est supérieur à 10 et la probabilité de succès est proche de 0,5. Lorsque la probabilité de succès est proche de 0 ou 1, un plus grand échantillon de Bernoulli est généralement requis.

Nous avons défini les effectifs d'échantillons pour les deux paires sur une valeur de n , où $n = 10, 15, 20, 30, \dots, 100$. Nous avons limité l'étude à des plans symétriques ($n_1 = n_2 = n$) sans en réduire le caractère de généralité, étant donné que les deux règles dépendent du nombre observé de succès et de défaillances, lequel peut être contrôlé par l'effectif des échantillons et la proportion de succès.

Pour estimer le niveau de confiance réel de l'intervalle de confiance pour la différence entre les deux proportions de population (appelé niveau de confiance simulée), nous avons calculé la proportion des 10 000 intervalles contenant la différence réelle entre les deux proportions. La probabilité de couverture cible dans chaque expérience était de 0,95. En outre, nous avons déterminé le pourcentage des 10 000 échantillons pour lesquels les conditions dans les deux règles étaient satisfaites.

Remarque : pour certains petits échantillons, l'erreur type estimée pour la différence entre les proportions était de 0. Nous avons considéré ces échantillons "dégénérés" et les avons supprimés de l'expérience. En conséquence, dans certains cas, le nombre de répliques d'échantillon était légèrement inférieur à 10 000.

Les résultats sont affichés dans les tableaux 1-11 et représentés graphiquement dans la figure 1 ci-dessous.

Tableau 1 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 10$. La probabilité de couverture cible est de 0,95.

n = 10							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	8,00	2,00	0,907	0,0	99,1
	p_2	0,90	9,00	1,00			
B	p_1	0,40	4,00	6,00	0,928	4,4	63,0
	p_2	0,55	5,50	4,50			
C	p_1	0,45	4,50	5,50	0,919	0,0	48,3
	p_2	0,90	9,00	1,00			

Tableau 2 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 15$. La probabilité de couverture cible est de 0,95.

n = 15							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	12,00	3,00	0,938	0,2	100,0
	p_2	0,90	13,50	1,50			

n = 15							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
B	p_1	0,40	6,00	9,00	0,914	65,0	97,3
	p_2	0,55	8,25	6,75			
C	p_1	0,45	6,75	8,25	0,93	1,2	86,9
	p_2	0,90	13,50	1,50			

Tableau 3 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour n = 20. La probabilité de couverture cible est de 0,95.

n = 20							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	16,00	4,00	0,942	1,5	100,0
	p_2	0,90	18,00	2,00			
B	p_1	0,40	8,00	12,00	0,943	92,8	99,9
	p_2	0,55	11,00	9,00			
C	p_1	0,45	9,00	11,00	0,934	4,1	98,2
	p_2	0,90	18,00	2,00			

Tableau 4 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour n = 30. La probabilité de couverture cible est de 0,95.

n = 30							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	24,00	6,00	0,941	4,3	100,0
	p_2	0,90	27,00	3,00			
B	p_1	0,40	12,00	18,00	0,944	99,7	100,0
	p_2	0,55	16,50	13,50			
C	p_1	0,45	13,50	16,50	0,938	7,2	100,0
	p_2	0,90	27,00	3,00			

Tableau 5 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 40$. La probabilité de couverture cible est de 0,95.

n = 40							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p₁	0,80	32,00	8,00	0,941	35,1	100,0
	p₂	0,90	36,00	4,00			
B	p₁	0,40	16,00	24,00	0,945	100,0	100,0
	p₂	0,55	22,00	18,00			
C	p₁	0,45	18,00	22,00	0,945	37,7	100,0
	p₂	0,90	36,00	4,00			

Tableau 6 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 50$. La probabilité de couverture cible est de 0,95.

n = 50							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p₁	0,80	40,00	10,00	0,942	36,4	100,0
	p₂	0,90	45,00	5,00			
B	p₁	0,40	20,00	30,00	0,944	100,0	100,0
	p₂	0,55	27,50	22,50			
C	p₁	0,45	22,50	27,50	0,935	38,3	100,0
	p₂	0,90	45,00	5,00			

Tableau 7 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 60$. La probabilité de couverture cible est de 0,95.

n = 60							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p₁	0,80	48,00	12,00	0,947	72,8	100,0
	p₂	0,90	54,00	6,00			
B	p₁	0,40	24,00	36,00	0,947	100,0	100,0
	p₂	0,55	33,00	27,00			

n = 60							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
C	p_1	0,45	27,00	33,00	0,949	73,1	100,0
	p_2	0,90	54,00	6,00			

Tableau 8 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour n = 70. La probabilité de couverture cible est de 0,95.

n = 70							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	56,00	14,00	0,939	71,7	100,0
	p_2	0,90	63,00	7,00			
B	p_1	0,40	28,00	42,00	0,945	100,0	100,0
	p_2	0,55	38,50	31,50			
C	p_1	0,45	31,50	38,50	0,944	71,8	100,0
	p_2	0,90	63,00	7,00			

Tableau 9 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour n = 80. La probabilité de couverture cible est de 0,95.

n = 80							
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2	
A	p_1	0,80	64,00	16,00	0,947	91,3	100,0
	p_2	0,90	72,00	8,00			
B	p_1	0,40	32,00	48,00	0,947	100,0	100,0
	p_2	0,55	44,00	36,00			
C	p_1	0,45	36,00	44,00	0,948	91,3	100,0
	p_2	0,90	72,00	8,00			

Tableau 10 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 90$. La probabilité de couverture cible est de 0,95.

n = 90						
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2
A	p₁	0,80	72,00	18,00	0,947	95,18
	p₂	0,90	81,00	9,00		
B	p₁	0,40	36,00	54,00	0,951	100,0
	p₂	0,55	49,50	40,50		
C	p₁	0,45	40,50	49,50	0,945	95,2
	p₂	0,90	81,00	9,00		

Tableau 11 Probabilité de couverture simulée et pourcentage d'échantillons satisfaisant les règles 1 et 2 pour $n = 100$. La probabilité de couverture cible est de 0,95.

n = 100						
Catégorie	Proportion (p)	np	n(1 - p)	Probabilité de couverture	% d'échantillons satisfaisant la règle 1	% d'échantillons satisfaisant la règle 2
A	p₁	0,80	80,00	20,00	0,952	97,7
	p₂	0,90	90,00	10,00		
B	p₁	0,40	40,00	60,00	0,945	100,0
	p₂	0,55	55,00	45,00		
C	p₁	0,45	45,00	55,00	0,948	97,7
	p₂	0,90	90,00	10,00		

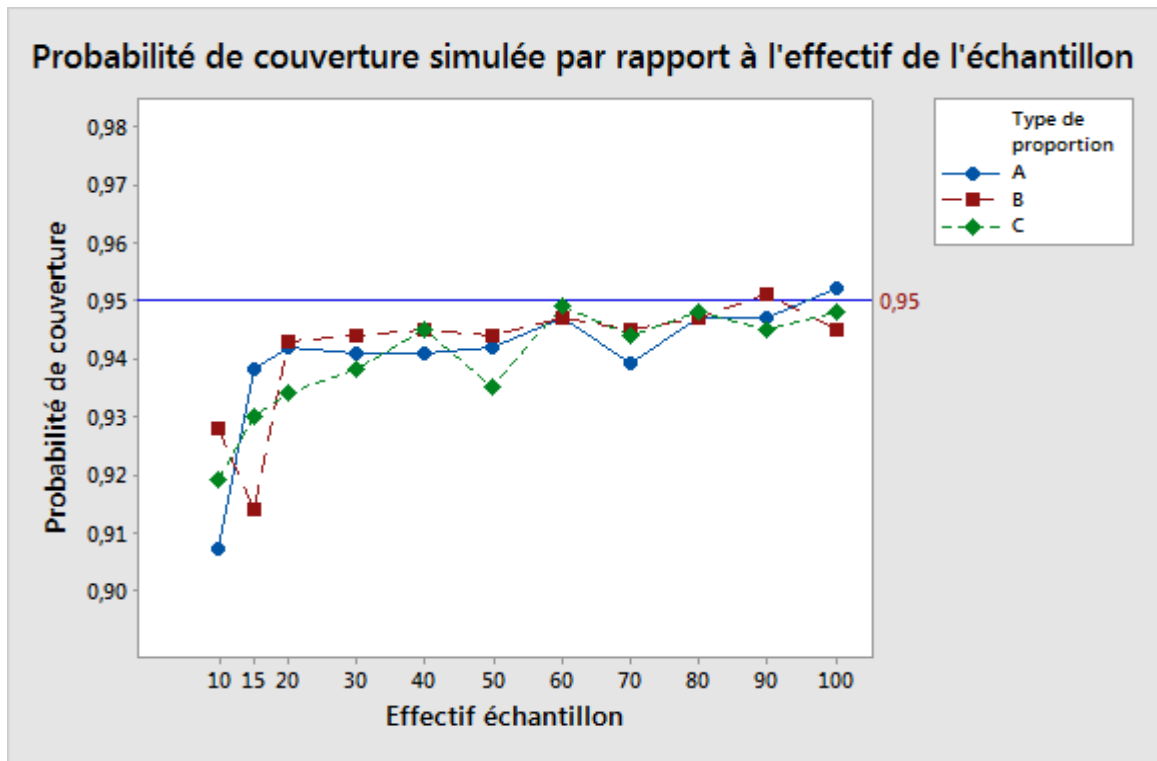


Figure 1 Probabilités de couverture simulée en fonction de l'effectif d'échantillon pour chaque catégorie de population de Bernoulli.

Les résultats des tableaux 1-11 et la figure 1 montrent que les échantillons générés à partir des populations de Bernoulli appartenant à la catégorie B (les deux proportions sont proches de 0,5) produisent généralement une probabilité de couverture simulée plus stable et plus proche de la couverture cible de 0,95. Les nombres de succès et de défaillances attendus dans les deux populations sont supérieurs dans cette catégorie, même lorsque les échantillons sont petits.

A l'inverse, pour les échantillons générés à partir de paires de populations de Bernoulli de catégorie A (lorsque les deux proportions sont proches de 1,0) ou C (lorsqu'une proportion est proche de 1,0 et l'autre de 0), la probabilité de couverture simulée est hors cible dans les petits échantillons, excepté lorsque le nombre de succès attendu (np) ou le nombre de défaillances attendu ($n(1-p)$) sont suffisamment grands.

Par exemple, examinons les échantillons générés à partir de populations de Bernoulli appartenant à la catégorie A, lorsque $n = 15$. Pour chaque population, le nombre de succès attendu est de 12,0 et 13,5, et le nombre de défaillances attendu de 3,0 et 1,5, respectivement. Bien que le nombre de défaillances attendu soit inférieur à 5 pour les deux populations, la probabilité de couverture simulée est d'environ 0,94. C'est sur la base de résultats comme ceux-ci que nous avons créé la règle 2, qui exige uniquement que le nombre de succès ou de défaillances prévu soit supérieur ou égal à 5 pour chaque échantillon.

Afin d'évaluer plus en détail l'efficacité des règles 1 et 2 pour évaluer l'approximation de l'intervalle de confiance, nous avons créé des graphiques représentant le pourcentage d'échantillons qui satisfont la règle 1 et le pourcentage d'échantillons qui satisfont la règle 2

en fonction des probabilités de couverture simulée obtenues dans les expériences. Ces graphiques sont visibles dans la figure 2.

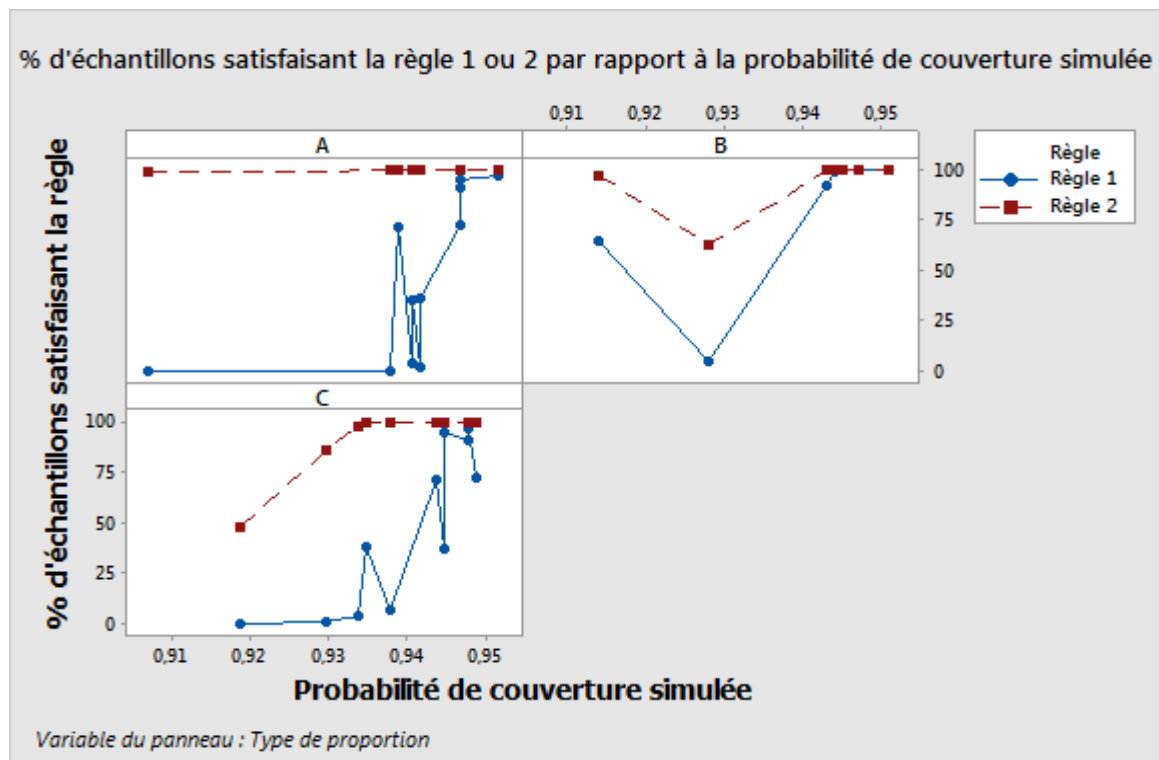


Figure 2 Pourcentages d'échantillons satisfaisant les règles 1 et 2 en fonction de la probabilité de couverture simulée pour chaque catégorie de population de Bernoulli.

Les graphiques montrent que, lorsque la probabilité de couverture simulée est proche de la couverture cible de 0,95, le pourcentage d'échantillons qui satisfont aux exigences de chaque règle est généralement proche de 100 %. Pour les échantillons générés à partir de populations de Bernoulli appartenant aux catégories A et C, la règle 1 est stricte lorsque les échantillons sont petits, comme le montre le très faible pourcentage d'échantillons satisfaisant la règle, même si la probabilité de couverture simulée est proche de la cible. Par exemple, lorsque $n = 20$ et que les échantillons sont issus de populations de Bernoulli appartenant à la catégorie A, la probabilité de couverture simulée est de 0,942 (voir le tableau 3). Toutefois, la proportion d'échantillons qui satisfont la règle est proche de 0 (0,015) (voir la figure 2). Dans ces cas, la règle risque donc être trop prudente.

En revanche, la règle 2 est moins stricte pour les petits échantillons générés à partir de populations de Bernoulli appartenant à la catégorie A. Par exemple, comme l'indique le tableau 1, lorsque $n = 10$ et que les échantillons sont issus de populations de Bernoulli appartenant à la catégorie A, la probabilité de couverture simulée est de 0,907 et 99,1 % des échantillons satisfont la règle.

En conclusion, la règle 1 a tendance à être trop prudente lorsque les échantillons sont petits. La règle 2 est moins prudente et peut être préférable lorsque les effectifs d'échantillons sont faibles. Cela dit, la règle 1 est connue et acceptée. Bien que la règle 2 présente un potentiel intéressant, elle peut être trop libérale dans certains cas, comme nous l'avons constaté précédemment. Il serait envisageable de combiner les deux règles pour bénéficier de leurs

avantages respectifs ; toutefois, cette approche doit être étudiée de façon plus approfondie avant de pouvoir être appliquée.

Annexe E : comparaison de la puissance réelle et de la puissance théorique

Simulation E1 : estimation de la puissance réelle à l'aide du test exact de Fisher

Nous avons mis en place une simulation pour comparer les niveaux de puissance réelle estimés (appelés niveaux de puissance simulée) du test exact de Fisher aux niveaux de puissance théorique associés au test d'approximation selon la loi normale (appelés niveaux de puissance par approximation). Dans chaque expérience, nous avons généré 10 000 paires d'échantillons à partir de paires de populations de Bernoulli. Pour chaque paire d'échantillons, les proportions ont été choisies pour que leur différence soit $p_1 - p_2 = -0,20$.

- **Proportions A : p_1 et p_2 sont proches de 1,0 (ou de 0).** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,70$ et $p_2 = 0,90$.
- **Proportions B : p_1 et p_2 sont proches de 0,5.** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,40$ et $p_2 = 0,60$.
- **Proportions C : p_1 est proche de 0,5 et p_2 est proche de 1,0.** Pour représenter cette paire de populations de Bernoulli dans la simulation, nous avons utilisé $p_1 = 0,55$ et $p_2 = 0,75$.

Nous avons défini les effectifs d'échantillons pour les deux paires sur une valeur de n , où $n = 10, 15, 20, 30, \dots, 100$. Nous avons limité l'étude à des plans symétriques ($n_1 = n_2 = n$), car l'on suppose généralement que les deux échantillons ont le même effectif. Nous avons calculé l'effectif d'échantillon commun nécessaire pour détecter une différence pratique importante avec une puissance déterminée.

Pour estimer la puissance réelle du test exact de Fisher à partir des résultats de chaque simulation, nous avons calculé la fraction des 10 000 paires d'échantillons pour lesquels le test bilatéral était significatif au seuil de signification cible, $\alpha = 0,05$. Pour effectuer la comparaison, nous avons ensuite calculé les niveaux de puissance théorique correspondants pour le test d'approximation selon la loi normale. Les résultats sont présentés dans le tableau 12 ci-dessous.

Tableau 12 Comparaison des niveaux de puissance simulée du test exact de Fisher et des niveaux de puissance par approximation pour les trois catégories de populations de Bernoulli. Le seuil de signification cible est de $\alpha = 0,05$.

n	Proportions A		Proportions B		Proportions C	
	<p>p₁ = 0,70 p₂ = 0,90</p>		<p>p₁ = 0,40 p₂ = 0,60</p>		<p>p₁ = 0,55 p₂ = 0,75</p>	
	Puissance simulée	Puissance par app.	Puissance simulée	Puissance par app.	Puissance simulée	Puissance par app.
10	0,063	0,193	0,056	0,14	0,056	0,149
15	0,151	0,271	0,097	0,19	0,101	0,204
20	0,244	0,348	0,146	0,24	0,183	0,259
30	0,37	0,49	0,256	0,338	0,272	0,366
40	0,534	0,612	0,371	0,431	0,381	0,466
50	0,641	0,711	0,477	0,516	0,491	0,556
60	0,726	0,789	0,536	0,593	0,56	0,635
70	0,814	0,849	0,61	0,661	0,649	0,703
80	0,87	0,893	0,66	0,72	0,716	0,76
90	0,907	0,925	0,716	0,77	0,772	0,808
100	0,939	0,948	0,792	0,812	0,812	0,848

Les résultats du tableau 12 montrent que la puissance obtenue par approximation a tendance à être supérieure à la puissance simulée pour les trois catégories de populations de Bernoulli (A, B et C). Par exemple, pour les proportions de catégorie A, l'effectif d'échantillon réellement nécessaire pour détecter une différence absolue de -0,20 avec un niveau de puissance par approximation de 0,91 est d'environ 90. Par comparaison, l'effectif d'échantillon correspondant estimé à l'aide de la fonction puissance théorique par approximation est d'environ 85. Par conséquent, l'effectif d'échantillon estimé à l'aide de la fonction puissance par approximation est souvent légèrement inférieur à l'effectif d'échantillon réellement nécessaire pour atteindre un niveau de puissance donné.

Cette relation est clairement visible lorsque les résultats sont affichés sous forme de courbes de puissance, comme le montre la figure 3 ci-dessous.

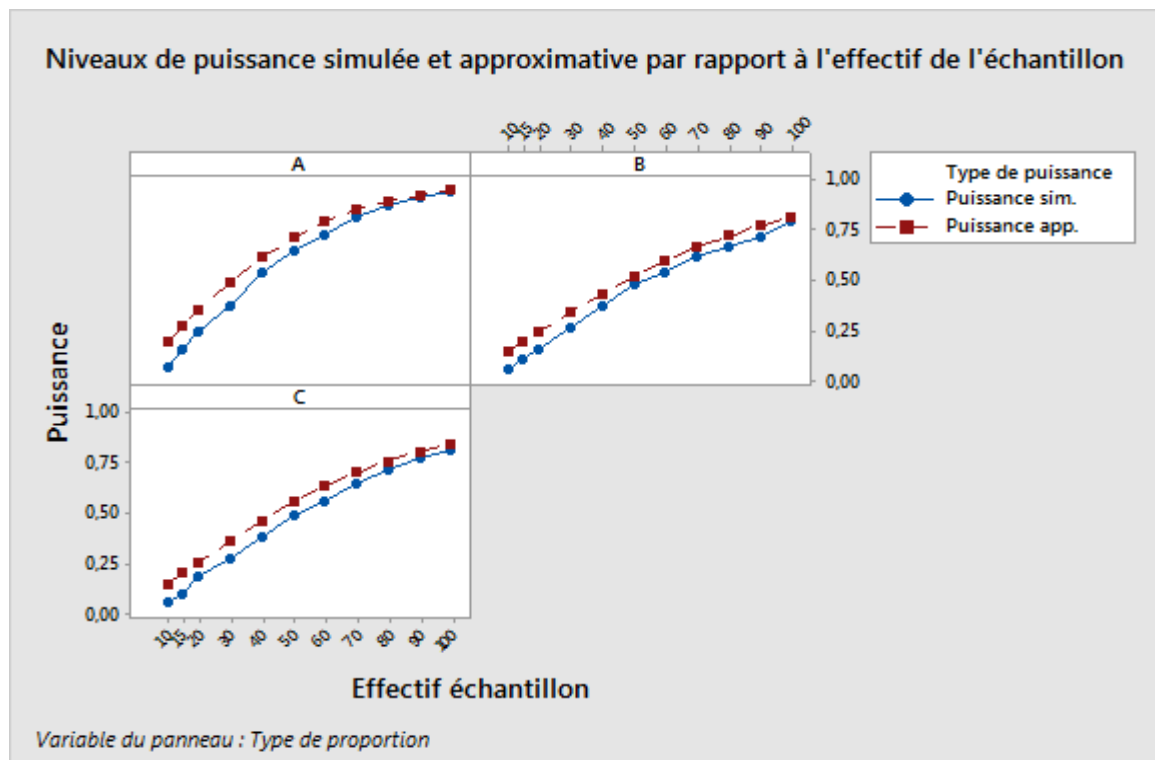


Figure 3 Graphiques des niveaux de puissance simulée et par approximation du test bilatéral pour la comparaison de deux proportions. Le graphique représente les courbes de puissance en fonction de l'effectif d'échantillon sur des panneaux correspondant à chaque catégorie de population de Bernoulli.

Bien que les courbes de puissance simulée soient inférieures aux courbes de puissance par approximation pour les trois catégories de populations de Bernoulli (A, B et C), l'importance de la différence entre les courbes dépend des proportions réelles des populations de Bernoulli à partir desquelles les échantillons sont générés. Par exemple, lorsque les deux proportions sont proches de 0,5 (catégorie B), les deux niveaux de puissance sont généralement proches. Toutefois, l'écart entre les deux courbes de puissance est plus visible dans de petits échantillons pour les proportions associées aux populations de catégorie A et C.

Ces résultats montrent que, de manière générale, la fonction puissance théorique du test d'approximation selon la loi normale et la fonction puissance simulée du test exact de Fisher sont approximativement égales. Par conséquent, l'Assistant utilise la fonction puissance théorique du test d'approximation selon la loi normale pour calculer les effectifs d'échantillons avant d'effectuer le test exact de Fisher. Toutefois, les effectifs d'échantillons calculés à l'aide de la fonction puissance par approximation peuvent être légèrement inférieurs aux effectifs d'échantillons réellement nécessaires pour atteindre une puissance donnée permettant de détecter une différence entre les deux proportions (pourcentages de défectueux).

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See minitab.com/legal/trademarks for more information.