Ce livre blanc fait partie d'une série de documents qui expliquent les recherches menées par les statisticiens de Minitab pour développer les méthodes et les outils de vérification des données utilisés dans l'Assistant de Minitab Statistical Software.

Test de pourcentage de défectueux à 1 échantillon

Généralités

Un test à 1 proportion permet de déterminer si une proportion diffère d'une valeur cible. 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 d'un pourcentage cible de défectueux.

L'Assistant de Minitab est doté d'un test de pourcentage de défectueux à 1 échantillon. Les données collectées pour le test représentent le nombre de défectueux dans un échantillon. Ce nombre est 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 et les intervalles de confiance ; aussi, le taux d'erreur de 1ère espèce réel est a priori proche du seuil de signification (alpha) spécifié pour le test et aucune étude plus approfondie n'est requise. En revanche, l'analyse de puissance et d'effectif d'échantillon du test de pourcentage de défectueux à 1 échantillon s'appuie sur une approximation ; il est donc nécessaire d'en évaluer l'exactitude.

Dans cet article, nous étudions la méthode utilisée pour évaluer la puissance et l'effectif d'échantillon pour le test de pourcentage de défectueux à 1 échantillon, et comparons la puissance théorique obtenue par approximation à la puissance réelle du test exact.

Nous expliquons également comment nous avons défini les conditions vous permettant de déterminer si l'effectif d'échantillon est suffisamment grand pour détecter si le pourcentage d'éléments défectueux diffère de la valeur cible. L'Assistant teste automatiquement l'effectif de l'échantillon et présente les résultats dans le rapport.

Le test de pourcentage de défectueux à 1 échantillon dépend également d'autres hypothèses. Pour plus de détails, reportez-vous à l'Annexe A.

Méthode de test de pourcentage de défectueux à 1 échantillon

Efficacité de la fonction puissance théorique

L'Assistant effectue le test d'hypothèse pour déterminer la proportion (pourcentage de défectueux) d'une seule population de Bernoulli à l'aide de méthodes exactes (rapport de vraisemblance). 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 d'approximation selon la loi normale correspondant.

Objectif

Nous souhaitions déterminer si nous pouvions utiliser la fonction puissance théorique issue du test d'approximation selon la loi normale pour évaluer la puissance et l'effectif d'échantillon nécessaires pour le test de pourcentage de défectueux à 1 échantillon dans l'Assistant. Pour ce faire, il nous fallait déterminer si cette fonction puissance théorique reflétait la puissance réelle du test exact (rapport de vraisemblance).

Méthode

La statistique de test, la valeur de p et l'intervalle de confiance du test exact (rapport de vraisemblance) sont définis dans l'Annexe B. La fonction puissance théorique issue de l'approximation selon la loi normale est définie dans l'Annexe C. Partant de ces définitions, nous avons réalisé des simulations afin d'estimer les niveaux de puissance réels (que nous appelons niveaux de puissance simulés) obtenus avec le test exact.

Pour ce faire, nous avons généré des échantillons aléatoires d'effectifs variés à partir de plusieurs populations de Bernoulli. Pour chaque population de Bernoulli, nous avons effectué le test exact sur chacune des 10 000 répliques d'échantillons. 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 échantillons pour lesquels 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 D.

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 (rapport de vraisemblance) 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'échantillon garantissant que le test exact soit suffisamment puissant pour détecter des différences pratiques importantes dans le pourcentage de défectueux.

Vérification des données

Effectif d'échantillon

Par définition, un test d'hypothèse vise à collecter des preuves permettant de rejeter l'hypothèse nulle de "non-différence". Lorsque l'échantillon est trop petit, la puissance du test peut ne pas être adaptée pour détecter une différence existante, ce qui entraîne une erreur de 2ème espèce. Il est donc essentiel de s'assurer que les effectifs d'échantillon 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'échantillon sont suffisamment grands pour que le test détecte des différences pratiques avec une probabilité élevée. Même si la planification d'effectif d'échantillon vise à garantir que les effectifs d'échantillon 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 1 échantillon 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 (reportez-vous à la section Méthode de test de pourcentage de défectueux à 1 échantillon ci-dessus). Avec un pourcentage cible de défectueux donné, la fonction puissance théorique dépend de l'effectif d'échantillon et de la différence que vous souhaitez détecter.

Les résultats

Lorsque les données ne fournissent pas suffisamment de preuves invalidant l'hypothèse nulle, l'Assistant calcule 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 présentant un intérêt particulier, l'Assistant calcule les effectifs d'échantillon 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 1 échantillon affiche les indicateurs d'état suivants lors du test de la puissance et de l'effectif d'échantillon :

Etat	Condition
	Le test détecte une différence entre le pourcentage de défectueux et la valeur cible, la puissance n'est donc pas un problème. OU La puissance est suffisante. Le test n'a pas détecté de différence avec la valeur cible, mais l'échantillon est suffisamment grand pour fournir une probabilité d'au moins 90 % de détecter la différence donnée (puissance ≥ 0,90).
	Il se peut que la puissance soit suffisante. Le test n'a pas détecté de différence avec la valeur cible, mais l'échantillon est suffisamment grand pour fournir une probabilité de 80 % à 90 % de détecter la différence donnée (0,80 ≤ puissance < 0,90). L'effectif d'échantillon nécessaire pour atteindre une puissance de 90 % est indiqué.
	Il se peut que la puissance ne soit pas suffisante. Le test n'a pas détecté de différence avec la valeur cible, et l'échantillon est suffisamment grand pour fournir une probabilité de 60 % à 80 % de détecter la différence donnée (0,60 ≤ puissance < 0,80). Les effectifs d'échantillon nécessaires pour atteindre une puissance de 80 % et de 90 % sont indiqués.
	La puissance n'est pas suffisante. Le test n'a pas détecté de différence avec la valeur cible, 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.
	Le test n'a pas détecté de différence avec la valeur cible. Vous n'avez pas indiqué de différence pratique à détecter. Par conséquent, le rapport indique les différences ayant 80 et 90 % de chances d'être détectées avec les effectifs d'échantillons et la valeur alpha utilisés.

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 pourcentage de défectueux à 1 échantillon

Le test de pourcentage de défectueux à 1 échantillon repose sur les hypothèses suivantes :

- Les données sont composées de n éléments distincts, chaque élément étant classé comme conforme ou non conforme.
- 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 (rapport de vraisemblance)

Supposons que nous observons un échantillon aléatoire $X_1, \dots, X_n$ issu d'une population de Bernoulli avec une probabilité de succès de $p = \Pr(X_i = 1) = 1 - \Pr(X_i = 0)$.

Les méthodes exactes pour la réalisation d'une inférence sur p sont décrites ci-dessous.

Formule B1 : test exact et valeur de p

Imaginons que l'on teste l'hypothèse nulle $H_0: p = p_0$ contre l'une de ces hypothèses alternatives : $H_A: p > p_0$, $H_A: p < p_0$ ou $H_A: p \neq p_0$.

Soit $X = \sum_{i=1}^{n} X_i$

Alors X est une variable aléatoire binomiale avec un nombre d'essais n et une probabilité de succès p.

Un test unilatéral fondé sur X est UPP (uniformément le plus puissant) et constitue un test de rapport de vraisemblance. Pour les tests bilatéraux, le test du rapport de vraisemblance s'appuie également sur X, et la statistique de test est

$$\Lambda(X) = \left(\frac{\hat{p}}{p_0}\right)^X \left(\frac{1-\hat{p}}{1-p_0}\right)^{n-X}$$

(voir Arnold, 1990).

Il est possible d'obtenir les valeurs de p des tests unilatéraux directement à partir de la loi de distribution exacte de X. Pour les tests bilatéraux, les valeurs de p sont calculées comme la probabilité, sous l'hypothèse nulle, d'observer un rapport de vraisemblance (ou un rapport de log vraisemblance) au moins aussi grand que celui réellement observé. Un algorithme de recherche numérique de zéro est généralement utilisé pour calculer cette probabilité.

Formule B2 : intervalle de confiance exact

Un intervalle de confiance exact bilatéral à $100(1-\alpha)\%$ pour p est égal à

$$\frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \leq p \leq \frac{\frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}$$

où x est le nombre de succès observé et $F_{v_1,v_2,\alpha}$ le point de percentile α supérieur de la loi F à v_1 et v_2 degrés de liberté (voir Casella et Berger, 1990). Par convention, nous définissons 0 comme limite inférieure si $x = 0$ et 1 comme limite supérieure si $x = n$.

Annexe C : fonction puissance théorique

Le calcul de la fonction puissance théorique du test exact est trop complexe à dériver. Par conséquent, nous en déterminons une approximation à l'aide de la fonction puissance théorique du test d'approximation selon la loi normale. Le test d'approximation s'appuie sur le fait que la variable aléatoire

$$Z = \frac{n^{1/2}(\hat{p} - p)}{(p(1-p)^{1/2}}$$

est asymptotiquement distribuée selon la loi normale standard. La fonction puissance théorique de ce test est connue et a fait l'objet de plusieurs études. Pour l'hypothèse alternative bilatérale, la fonction puissance est exprimée de la façon suivante :

$$\pi(n,\delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

où $p = \delta + p_o$, $\Phi(.)$ est la fonction de répartition de la loi normale standard et z_α est le percentile supérieur de la loi normale standard.

Pour l'hypothèse alternative unilatérale $H_A: p > p_0$, la fonction puissance peut être exprimée comme suit :

$$\pi(n,\delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

Avec l'hypothèse alternative unilatérale $H_A: p < p_0$, la fonction puissance peut être exprimée comme suit :

$$\pi(n,\delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

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

Simulation D1 : estimation de la puissance réelle à l'aide du test exact

Nous avons mis en place une simulation pour comparer les niveaux de puissance réelle estimée (appelés niveaux de puissance simulée) aux niveaux de puissance théorique associés à la fonction puissance du test d'approximation selon la loi normale (appelés niveaux de puissance par approximation). Dans chaque expérience, nous avons généré 10 000 échantillons, tous d'effectif n, issus de populations de Bernoulli avec une probabilité de succès p donnée. Nous avons pris en compte deux cas de probabilité de succès : (1) une probabilité de succès modérée, avec une valeur de p proche de 0,5 (plus précisément, $p = 0{,}45$) et (2) une probabilité de succès faible ou élevée, avec une valeur de p proche de 0 ou 1 (plus précisément, $p = 0{,}85$). Nous avons pris en compte ces deux cas, car l'approximation selon la loi normale de DeMoivre-Laplace de la loi binomiale (de laquelle le test d'approximation est dérivé) est exacte lorsque l'effectif d'échantillon de Bernoulli est supérieur à 10 et la probabilité de succès proche de 0,5. Cependant, avec des probabilités de succès plus faibles ou plus élevées, de plus grands échantillons de Bernoulli sont nécessaires afin que l'approximation soit exacte.

Dans chaque expérience, nous avons défini l'effectif d'échantillon sur une valeur unique de n, où $n = 10, 15, 20, 30, \ldots, 100$. Pour toutes les expériences, nous avons défini la différence à détecter $\delta = p - p_0$ sur 0,2 pour garantir que les valeurs de puissance obtenues ne soient ni trop faibles ni trop élevées, à mesure que l'effectif d'échantillon passe à 100. Pour estimer la puissance réelle du test à partir des résultats de chaque simulation, nous avons calculé la fraction des 10 000 répliques d'échantillons pour lesquelles le test exact était significatif à un seuil de signification cible $\alpha = 0{,}05$, pour des tests exacts unilatéraux et bilatéraux. Enfin, pour effectuer la comparaison, nous avons calculé les niveaux de puissance théorique correspondants pour le test d'approximation selon la loi normale. Les résultats sont présentés dans le tableau 1 ci-dessous.

Tableau 1 Niveaux de puissance simulée et par approximation des tests exacts unilatéraux et bilatéraux. Le seuil de signification cible est de $\alpha = 0{,}05$.

n	Test bilatéral				Test unilatéral			
	$p = 0{,}45$		$p = 0{,}85$		$p = 0{,}45$		$p = 0{,}85$	
	Puissance simulée	Puissance par app.	Puissance simulée	Puissance par app.	Puissance simulée	Puissance par app.	Puissance simulée	Puissance par app.
10	0,101	0,333	0,20	0,199	0,257	0,436	0,20	0,335
15	0,339	0,441	0,322	0,327	0,339	0,55	0,322	0,489
20	0,406	0,537	0,409	0,455	0,59	0,643	0,648	0,621
30	0,632	0,69	0,708	0,674	0,632	0,779	0,708	0,808
40	0,781	0,799	0,863	0,822	0,781	0,867	0,863	0,911
50	0,877	0,872	0,874	0,91	0,877	0,921	0,933	0,961
60	0,878	0,92	0,942	0,957	0,922	0,954	0,969	0,984
70	0,925	0,951	0,972	0,981	0,953	0,973	0,987	0,994
80	0,954	0,971	0,986	0,992	0,986	0,985	0,993	0,998
90	0,971	0,982	0,993	0,996	0,991	0,991	0,996	0,999
100	0,989	0,99	0,998	0,999	0,994	0,995	0,999	1,00

Les résultats indiquent que les niveaux de puissance simulée et par approximation sont généralement très proches. Cette concordance est clairement visible lorsque les résultats sont représentés graphiquement sous forme de courbes de puissance, comme dans les figures 1 et 2 ci-dessous.

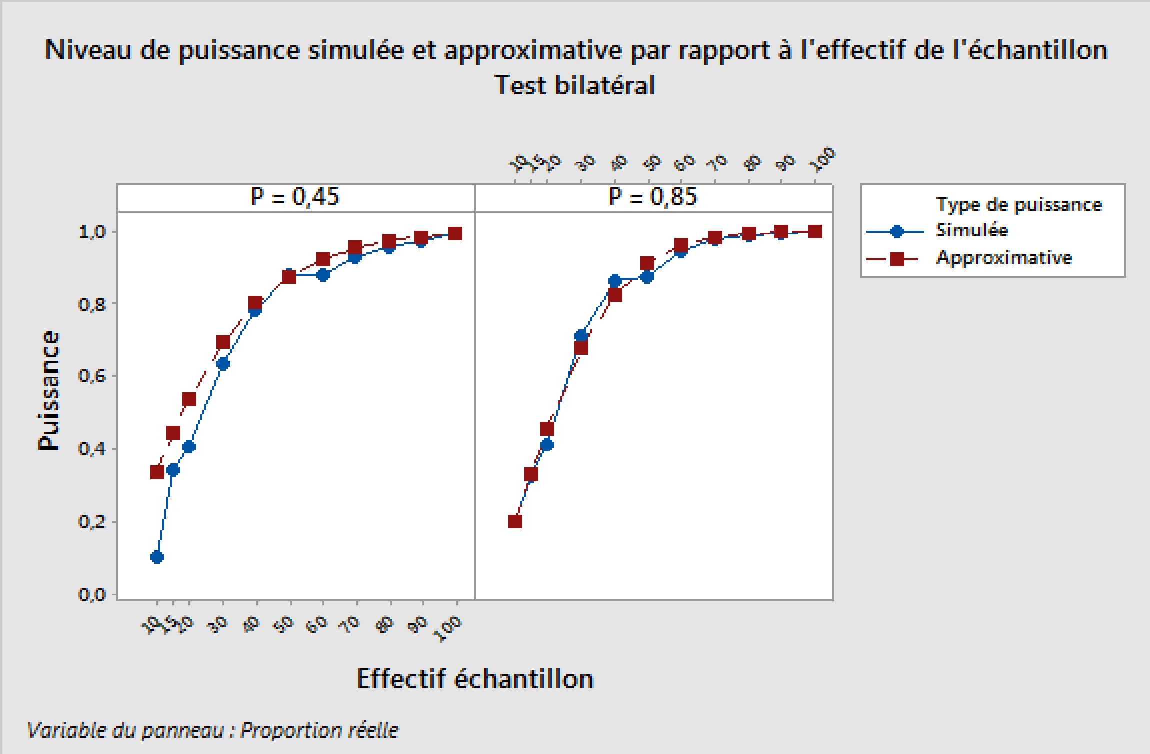

Figure 1 Graphiques des niveaux de puissance simulée et par approximation du test exact bilatéral en fonction de l'effectif d'échantillon.

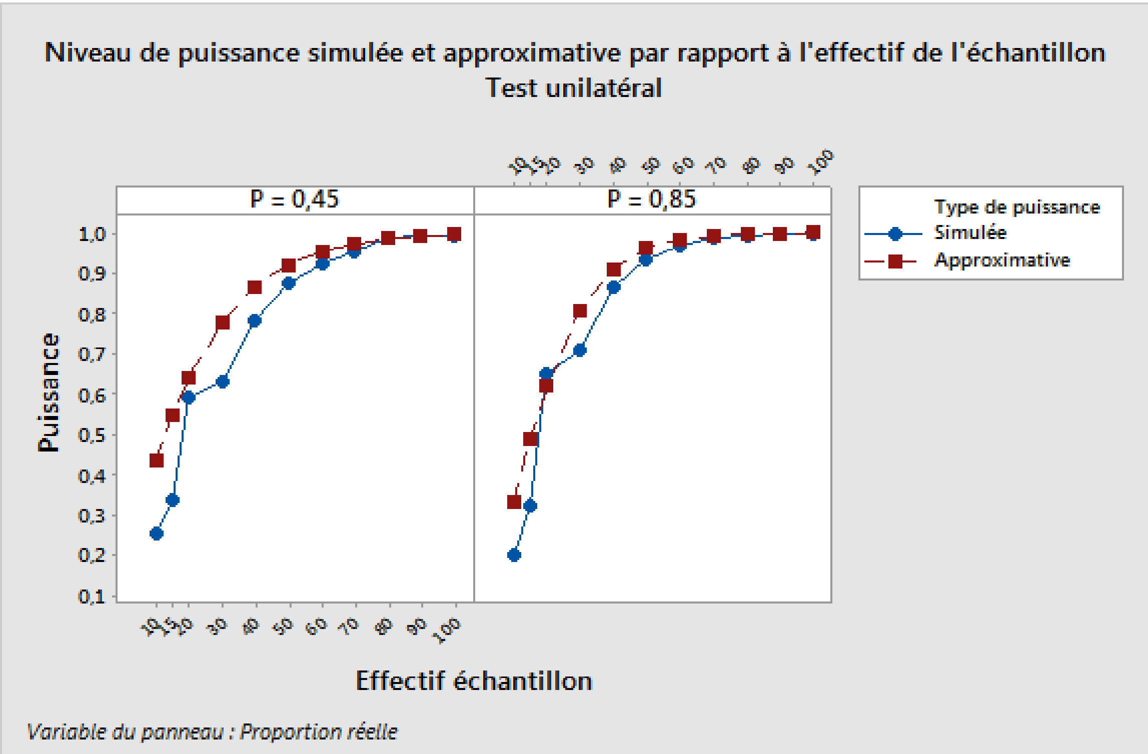

Figure 2 Graphiques des niveaux de puissance simulée et par approximation du test exact unilatéral en fonction de l'effectif d'échantillon.

Les deux courbes de puissance indiquées dans les graphiques des figures 1 et 2 sont proches, excepté dans certains cas où l'effectif d'échantillon est faible. La proximité des courbes indique que la fonction puissance par approximation correspond étroitement à la puissance simulée lorsque le test exact est appliqué dans la pratique. Par conséquent, nous pouvons utiliser la fonction puissance par approximation pour déterminer l'effectif d'échantillon.

Les figures 1 et 2 montrent également que les courbes de puissance théorique (par approximation) sont généralement plus élevées que celles de puissance simulée. Les courbes de puissance par approximation sont plus élevées, car les niveaux de puissance théorique sont calculés à partir d'une valeur exacte pour le seuil de signification cible (0,05). A l'inverse, le test exact a tendance à être prudent, en particulier avec de petits échantillons, et produit donc des seuils de signification réels inférieurs au seuil cible. Par conséquent, les niveaux de puissance simulée ont tendance à être plus faibles lorsque les effectifs d'échantillons sont petits.

En conclusion, nos simulations montrent que la fonction puissance théorique du test d'approximation selon la loi normale offre une approximation précise de la puissance du test exact (rapport de vraisemblance). Par conséquent, la fonction puissance théorique du test d'approximation selon la loi normale fournit une base solide pour déterminer les effectifs d'échantillon nécessaires afin de garantir que le test exact est suffisamment puissant pour détecter des différences importantes dans la pratique.