

Méthode de Bonett

SUR L'INTERVALLE DE CONFIANCE ROBUSTE DE BONETT POUR LE RAPPORT DES ECARTS TYPES

Senin J. Banga et Gregory D. Fox
18 juin 2013

ABSTRACT

Nous proposons une procédure alternative pour corriger une erreur légère dans l'intervalle de confiance (IC) de Bonett (2006) pour le rapport de deux écarts types. L'estimateur d'aplatissement regroupé de la statistique de test de Layard (1973), sur lequel est fondé l'intervalle de Bonett, n'est cohérent que lorsque les variances des populations sont égales. Nous délivrons donc un autre estimateur, cohérent aussi bien lorsque les variances des populations sont égales que lorsqu'elles sont inégales, et utilisons ce nouvel estimateur pour calculer un intervalle de confiance approprié. Des études par simulation indiquent que, de manière générale, ce nouvel IC est plus exact et plus précis que celui fondé sur le test W_{50} de Levene/Browne-Forsythe et sur le test L_{50} de Pan (1999). Conformément aux observations de Pan, nous constatons que les IC fondés sur le test W_{50} sont moins précis avec de petits échantillons, ce qui crée souvent des intervalles de largeur infinie. Les IC fondés sur le test L_{50} permettent d'obtenir de bons résultats avec des lois symétriques et presque symétriques, mais pas avec des populations asymétriques.

Termes d'index : homogénéité des variances, test de Levene, test de Brown-Forsythe, test de Layard, intervalle de confiance (IC) pour le rapport des variances

1. Introduction

Nous savons que le test F classique et les intervalles de confiance (IC) qui lui sont associés sont extrêmement sensibles à des écarts par rapport à la normalité : si sensibles en fait que le test F classique n'est pas adapté à la plupart des applications pratiques. C'est pourquoi de nombreuses autres solutions, plus robustes, ont été proposées. Parmi celles-ci, le test connu sous le nom de "test W50" est souvent privilégié, car il offre des propriétés très satisfaisantes concernant le taux d'erreur de 1ère espèce, tout en étant facile à calculer et à interpréter. (Pour des analyses comparatives, voir Conover et al. (1981), Balakrishnan et Ma (1990), et Lim et Loh (1996).) Le test W_{50} repose sur une procédure initialement proposée par Levene (1960) et enrichie par Brown et Forsythe (1974). Le test W_{50} a été largement adopté et il est

aujourd'hui disponible dans la plupart des logiciels statistiques de référence, tels que Minitab Statistical Software, SAS, R et JMP.

Cependant, les propriétés du test W_{50} ne sont pas aussi remarquables pour le taux d'erreur de 2ème espèce que pour le taux d'erreur de 1ère espèce. Pan (1999) démontre que pour certaines lois de distribution (notamment la loi normale), la puissance du test W_{50} dans des problèmes à 2 échantillons présente une borne supérieure pouvant être bien inférieure à 1. En outre, la valeur de la différence entre les variances des populations n'a pas d'incidences sur cette borne supérieure. Cette carence s'étend naturellement aux IC calculés à partir du test W_{50} . Pan démontre qu'il existe une probabilité non négligeable qu'un IC calculé pour le rapport des variances de populations à partir du test W_{50} soit infini $(0, +\infty)$, et donc dénué d'intérêt. L'observation de Pan est confirmée par les résultats de nos propres simulations, présentés plus bas dans cet article.

Pan propose une procédure alternative, appelée L_{50} , pour pallier les limites de la procédure W_{50} . S'appuyant sur les résultats de simulations, Pan conclut que le test L_{50} est plus puissant que le test W_{50} , mais tout aussi robuste. En outre, il partage les propriétés asymptotiques souhaitées. Toutefois, les échantillons utilisés par Pan pour ses simulations étaient issus de populations obéissant à des lois (modérément) asymétriques à queues légères ou lourdes. Or, l'impact potentiel de l'asymétrie sur les résultats du test L_{50} avec de petits échantillons n'a pas été spécifiquement examiné.

Pan avance également que la procédure L_{50} est aussi puissante que d'autres procédures notoirement robustes, comme le test des rangs modifié de Fligner-Killeen et le test adaptatif de Hall-Padmanabhan. En pratique, cependant, le test des rangs modifié de Fligner-Killeen et le test adaptatif de Hall-Padmanabhan, sont moins pratiques que les tests L_{50} et W_{50} , car ils demandent de nombreux et lourds calculs.

Bonett (2006) a récemment proposé une autre procédure d'IC, fondée sur la version pour deux échantillons du test de l'homogénéité des variances de Layard (1973). Bonett inclut plusieurs ajustements visant à améliorer les performances de la procédure de Layard avec de petits échantillons. Par exemple, il propose un estimateur d'aplatissement regroupé qui est asymptotiquement équivalent à celui de Layard, mais qui limite le biais dû aux petits échantillons.

Malheureusement, ni l'estimateur d'aplatissement regroupé initial de Layard, ni la solution de remplacement proposée par Bonett ne sont cohérents lorsque les variances des populations sont inégales. Par conséquent, les intervalles proposés par Bonett (2006) ne sont pas vraiment des IC au sens propre du terme ; il s'agirait plutôt d'intervalles d'acceptation pour le test d'égalité des variances. Ainsi, en soustrayant à 1 les probabilités de couverture simulée obtenues par Bonett (2006), nous obtenons les taux d'erreur de 1ère espèce du test d'égalité des variances. La comparaison de ces taux d'erreur de 1ère espèce à ceux du test initial de Layard confirme que les ajustements de Bonett améliorent les performances du test de Layard pour les petits échantillons. Cela dit, l'IC pour le rapport des variances proposé par Bonett doit être réexaminé.

En outre, notez que Bonett compare les intervalles qu'il propose aux IC obtenus avec le test F par approximation de Shoemaker (2003). Or, l'IC pour le rapport des variances associé au test de Shoemaker (brièvement décrit à la page 106 de l'article de Shoemaker) repose également sur l'estimateur d'aplatissement regroupé de Layard. Par conséquent, les IC

calculés dans la section 7 de l'article de Shoemaker s'apparentent eux aussi davantage à des intervalles d'acceptation pour le test d'égalité des variances. Malgré ces erreurs, les résultats de la simulation effectuée par Bonett montrent que son ajustement améliore les performances du test de l'égalité des variances de Layard pour les petits échantillons, et que le test d'égalité des variances qu'il obtient offre de meilleurs résultats que celui de Shoemaker.

Dans cette étude, nous corrigeons l'erreur de Bonett (2006) en étendant la version pour deux échantillons du test de Layard, qui teste les hypothèses nulles sur le rapport des variances ou des écarts types. Pour cela, nous proposons d'utiliser un estimateur d'aplatissement regroupé qui soit cohérent pour tout rapport hypothétisé donné. Nous inversons ensuite la statistique du test pour obtenir l'IC du rapport. Enfin, nous réalisons des études par simulation pour évaluer les propriétés de robustesse du nouvel IC dans des plans à faibles échantillons. En outre, nous comparons les performances du nouvel IC à celles des IC associés au test F classique, au test W_{50} et au test L_{50} pour de petits échantillons.

2. Test de Layard et extension

Soit $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$ k échantillons indépendants. Chaque échantillon est indépendant et distribué de façon identique avec une moyenne $E(Y_{ij}) = \mu_i$ et une variance $\text{Var}(Y_{ij}) = \sigma_i^2 > 0$. Supposons en outre que les échantillons proviennent de populations présentant un aplatissement commun $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$. Nous observons que Layard utilise l'excès d'aplatissement $\gamma_e = \gamma - 3$.

Soit \bar{Y}_i et S_i la moyenne et l'écart type de l'échantillon i , respectivement. De même, soit $\tau^2 = 2 + (1 - 1/\bar{n})\gamma_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$, où $\bar{n} = \sum n_i/k$. Comme le signale Layard (1973), pour de grands échantillons, $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$.

Pour tester l'hypothèse nulle d'égalité des variances, Layard effectue une transformation orthogonale sur le vecteur dont les composantes $Z_i = (n_i - 1)^{1/2} \ln S_i^2 / \tau$ sont distribuées asymptotiquement selon la loi normale standard sous l'hypothèse nulle. Il utilise ensuite la propriété de conservation de la distance des transformations orthogonales pour montrer que la statistique de test S' (ci-dessous) est distribuée asymptotiquement comme une loi du Khi deux à $k - 1$ degrés de liberté sous l'hypothèse nulle d'égalité des variances :

$$S' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \frac{\sum_{i=1}^k (n_i - 1) \ln S_i^2}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

Généralement, $Z_i = (n_i - 1)^{1/2} (\ln S_i^2 - \ln \sigma_i^2) / \tau$ est distribué asymptotiquement selon la loi normale standard. Par conséquent, il est possible d'appliquer les techniques de Layard pour calculer la statistique de test plus générale T'_k :

$$T'_k = \sum_{i=1}^k (n_i - 1) \frac{(\ln S_i^2 - \ln \sigma_i^2)^2}{\tau^2} - \left(\sum_{i=1}^k (n_i - 1) \frac{\ln S_i^2 - \ln \sigma_i^2}{\tau \sqrt{\sum_{i=1}^k (n_i - 1)}} \right)^2$$

T'_k est distribuée asymptotiquement selon une loi du Khi deux à $k - 1$ degrés de liberté sous les hypothèses nulle et alternative.

Il est possible d'exprimer T'_k de façon plus semblable à S' . En exprimant le terme mis au carré comme une somme double et en effectuant quelques manipulations algébriques, nous obtenons le résultat suivant :

$$T'_k = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \ln \sigma_i^2 - \frac{\sum_{i=1}^k (n_i - 1) (\ln S_i^2 - \ln \sigma_i^2)}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

Lorsque toutes les variances sont égales, alors $T'_k = S'$. Par conséquent, les statistiques S' et T'_k sont équivalentes pour le test de l'hypothèse nulle d'égalité des variances. Toutefois, il est également possible d'utiliser T'_k de façon plus générale pour tester toute hypothèse exprimée comme fonction des variances. Par exemple, il est possible d'utiliser T'_k pour tester toute hypothèse nulle de la forme $H_0: \sigma_i = \sigma_{0i}$ pour tout $\sigma_{0i} > 0, i = 1, \dots, k$ donné.

Etant donné que $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ est inconnu, un test s'appuyant sur S' ou T'_k requiert un estimateur de l'aplatissement regroupé des populations, γ . Par exemple, pour tester l'hypothèse nulle de l'homogénéité des variances, Layard propose l'estimateur d'aplatissement regroupé suivant :

$$\hat{\gamma} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right]^2} \sum_{i=1}^k n_i$$

Toutefois, Layard signale que la valeur de $\hat{\gamma}$ n'est pas toujours cohérente lorsque les variances sont inégales.

Dans le cas particulier des plans à 2 échantillons, il est possible d'évaluer la différence entre les écarts types en testant l'hypothèse nulle $H_0: \sigma_1/\sigma_2 = \rho_0$ pour un rapport hypothétisé $\rho_0 > 0$ donné. Toutefois, il est possible d'évaluer cette différence de façon plus directe en calculant l'IC pour le rapport des écarts types.

Si $\rho_0 = 1$, alors l'hypothèse nulle est celle de l'homogénéité des variances. Par conséquent, le test peut être fondé sur l'équivalence $T'_2 = S'$, en remplaçant γ par la version pour 2 échantillons de l'estimateur d'aplatissement de Layard dans l'expression de $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$, afin d'obtenir $\hat{\tau}^2$.

Toutefois, si $\rho_0 \neq 1$, alors le test doit être fondé sur T'_2 , et non sur S' . En outre, si $\rho_0 \neq 1$ alors l'estimateur d'aplatissement regroupé de Layard n'est pas nécessairement cohérent, et ne peut donc pas être utilisé pour estimer l'aplatissement regroupé des populations. Par conséquent, il nous faut utiliser un autre estimateur d'aplatissement regroupé, qui soit cohérent pour tout rapport hypothétisé $\rho_0 > 0$.

Nous proposons cet estimateur ci-après. Etant donné qu'il est fonction de ρ_0 , nous écrivons cet estimateur $\hat{\gamma}_p(\rho_0)$. En outre, nous définissons la statistique de test $T_2 = \tau^2 T'_2 / \hat{\tau}^2$, où $\hat{\tau}^2 = 2 + (1 - 1/\bar{n})(\hat{\gamma}_p(\rho_0) - 3)$. D'après le théorème de Slutsky, T_2 est distribué asymptotiquement comme une loi du Khi deux à 1 degré de liberté. Enfin, nous inversons T_2 pour obtenir les IC pour $\rho = \sigma_1/\sigma_2$.

3. IC pour le rapport des écarts types

La section précédente explique en quoi il est nécessaire d'utiliser un autre estimateur d'aplatissement lors du test d'hypothèses nulles fondées sur le rapport des variances ou des écarts types. Le résultat suivant fournit cet estimateur.

RESULTAT 1

Pour tout $\rho = \sigma_1/\sigma_2 > 0$ donné, l'aplatissement commun d'une population dans un modèle à 2 échantillons peut être évalué de façon cohérente à l'aide de l'estimateur regroupé

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

Ce résultat est démontré dans l'Annexe A.

Comme prévu, $\hat{\gamma}_P(1)$ est identique à l'estimateur d'aplatissement regroupé de Layard, $\hat{\gamma}$, puisque $\sigma_1/\sigma_2 = 1$ implique que les variances ou les écarts types sont égaux.

La statistique T'_2 , qui constitue la version pour 2 échantillons de la statistique générale T'_k , est exprimée comme suit :

$$T'_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) \tau^2}$$

où $\hat{\rho} = S_1/S_2$, $\rho = \sigma_1/\sigma_2$ et $\tau^2 = 2 + (1 - 1/\bar{n})\hat{\gamma}_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$.

Comme le signale Layard (1973), dans de grands échantillons, $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$. Bonett (2006) utilise une autre approximation, également adoptée par Shoemaker (2003), $\text{Var}((n_i - 1)^{1/2} \ln S_i^2) \cong \gamma - (n_i - 3)/n_i$. Ces approximations sont équivalentes dans de grands échantillons. Toutefois, Shoemaker signale que la seconde version est aussi utile lors de la réalisation de son test d'égalité des variances avec de petits échantillons. Cet ajustement permet de modifier la statistique T'_2 comme suit :

$$T'_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

où $g_i = (n_i - 3)/n_i$.

Ainsi, la statistique $T_2 = \tau^2 T'_2 / \hat{\tau}^2$ du test de l'hypothèse nulle $H_0: \rho = \rho_0$ peut être exprimée comme suit :

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{\gamma}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_0) - g_2}{n_2 - 1}}$$

Dans l'expression de T_2 , la racine carrée du dénominateur peut être considérée comme une estimation pour grands échantillons de l'erreur type de l'aplatissement regroupé.

En outre, dans l'expression de $\hat{\gamma}_P(1) \equiv \hat{\gamma}$, Bonett (2006) utilise les moyennes d'échantillons tronquées avec la proportion de troncature $1/[2(n_i - 4)^{1/2}]$. Nous effectuons le même ajustement sur l'estimateur d'aplatissement regroupé :

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

où m_i constitue la moyenne tronquée de l'échantillon i , avec la proportion de troncature $1/[2(n_i - 4)^{1/2}]$. Cette version de l'estimateur d'aplatissement regroupé et la version antérieure sont asymptotiquement équivalentes, car la moyenne tronquée m_i constitue un estimateur cohérent de la moyenne de la population μ_i . Toutefois, il se peut que la version alternative améliore les performances du test utilisant T_2 pour de petits échantillons.

Nous n'avons maintenant plus qu'à inverser la statistique de test T_2 pour dériver un IC par approximation pour le rapport des variances ou des écarts types. Cependant, nous allons d'abord revenir brièvement sur l'erreur présente dans le calcul des IC de Bonett (2006) pour le rapport des écarts types.

3.1 Intervalles de Bonett

Plutôt que d'inverser T_2 pour obtenir l'IC, Bonett (2006) inverse la statistique suivante :

$$T = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\hat{\gamma}_P(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(1) - g_2}{n_2 - 1}}$$

Par conséquent, l'intervalle résultant fournit simplement la zone d'acceptation pour le test d'égalité des variances. Cela est dû au fait que l'estimateur d'aplatissement regroupé $\hat{\rho}_P(1)$ n'est cohérent que lorsque les variances sont égales, c'est-à-dire lorsque le rapport hypothétisé est égal à 1. L'intervalle résultant est exprimé par Bonett (2006) sous la forme

$$\exp[\ln(c S_1^2/S_2^2) \pm z_{\alpha/2} es]$$

où

$$es^2 = \frac{\hat{\gamma}(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}(1) - g_2}{n_2 - 1}$$

La constante c est incluse afin d'ajuster légèrement l'échantillon pour réduire l'effet de queues inégales sur le calcul des probabilités d'erreur dans les plans non équilibrés. Cette constante est exprimée de la façon suivante :

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

La constante disparaît lorsque les plans sont équilibrés et son effet devient négligeable lorsque les effectifs d'échantillons augmentent.

Le tableau 1 illustre ce qui se passe lorsque l'on considère à tort les intervalles ci-dessus comme des IC. Ces résultats s'appuient sur une petite étude de simulation dans laquelle nous calculons les probabilités de couverture simulée d'intervalles de Bonett (2006). Pour les cas de variances égales (colonne de gauche), nous créons deux échantillons indépendants à partir de la loi normale standard. Pour les cas de variances inégales (colonne de droite), nous élevons les observations du deuxième échantillon à l'aide d'un facteur constant de 4. Les

probabilités de couverture estimée sont calculées sur la base de 100 000 répliques. La couverture nominale cible est de 0,95.

Tableau 1 : Effet de variances de populations inégales sur les IC de Bonett (2006) ($\alpha = 0,05$)

n_1, n_2	Probabilités de couverture simulée	
	Egalité des variances	Inégalité des variances
10, 10	0,963	0,972
50, 50	0,952	0,991
100, 100	0,952	0,994

Si les intervalles étaient calculés à l'aide d'un estimateur d'aplatissement regroupé cohérent, les probabilités de couverture devraient être identiques dans les deux cas. Toutefois, vous pouvez remarquer que les intervalles sont invariablement plus prudents lorsque les variances sont inégales. En outre, plus les effectifs d'échantillons augmentent, plus les probabilités de couverture se rapprochent de 1. Notez que les IC par approximation de Shoemaker (2003) produisent des résultats similaires.

3.2 Calculs pour l'IC

Examinons le cas du test de l'hypothèse nulle $H_0: \rho = \rho_0$ contre l'hypothèse alternative $H_A: \rho \neq \rho_0$, où $\rho = \sigma_1/\sigma_2$ et $\rho_0 > 0$, avec la statistique de test T_2 énoncée précédemment. Sous l'hypothèse nulle, la statistique de test

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1}}$$

est distribuée asymptotiquement comme une loi du Khi deux à 1 degré de liberté. Par conséquent, le test rejette l'hypothèse nulle au seuil de signification α si, et seulement si

$$(\ln \hat{\rho}^2 - \ln \rho_0^2)^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

où z_α désigne le $\alpha \times 100$ ème point de percentile supérieur de la loi normale standard. Notez que le $\alpha \times 100$ ème point de percentile supérieur de la loi du Khi deux à 1 degré de liberté, $\chi_{1,\alpha}^2$, satisfait la condition suivante : $\chi_{1,\alpha}^2 = z_{\alpha/2}^2$.

Les résultats des simulations effectuées par Bonett (2006) indiquent que l'ajustement fourni pour les petits échantillons en vue de réduire l'effet de queues inégales sur le calcul des probabilités d'erreur dans les plans non équilibrés permet d'obtenir de bons résultats. Par conséquent, nous avons effectué un ajustement similaire pour le test fondé sur la statistique T_2 . Une fois l'ajustement effectué, le test rejette l'hypothèse nulle si, et seulement si

$$(\ln \rho_0^2 - \ln(c\hat{\rho}^2))^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

où c représente la constante d'ajustement de Bonett, exprimée comme suit :

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Ainsi, il est possible d'exprimer un niveau de confiance approximatif de $(1 - \alpha)100$ pour cent pour $\rho = \sigma_1/\sigma_2$ avec T_2 de la façon suivante :

$$\left\{ \rho \in (0, \infty) : (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho) - g_2}{n_2 - 1} \right) \leq 0 \right\}$$

Notez que c n'a aucun effet dans les plans équilibrés et n'a qu'un effet négligeable dans les plans non équilibrés à grands échantillons.

Le résultat suivant fournit une autre façon d'exprimer la valeur de la confiance, qui nous permet d'en décrire plus facilement la nature. Dans cette expression, l'estimateur d'aplatissement regroupé est exprimé en fonction des aplatissements individuels des échantillons, obtenus comme suit :

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

RESULTAT 2

Il est possible d'exprimer un niveau de confiance approximatif de $(1 - \alpha)100$ pour cent pour $\rho = \sigma_1/\sigma_2$ avec T_2 de la façon suivante :

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty) : H(r^2) \leq 0\}$$

ou, de façon équivalente, le niveau de confiance défini pour $\rho^2 = \sigma_1^2/\sigma_2^2$ peut être exprimé comme suit :

$$c\hat{\rho}^2 \{r \in (0, \infty) : H(r) \leq 0\}$$

où

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 es^2(cx), x > 0$$

$$es^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{g_1}{n_1 - 1} + \frac{g_2}{n_2 - 1}, K = \frac{n_1 - 1}{n_2 - 1}$$

Ce résultat est démontré dans l'Annexe B.

Il est facile de vérifier que la fonction $H(x)$ est continue sur la droite des réels positifs, avec $H(0) = H(+\infty) = +\infty$ et $H(1) < 0$. Par conséquent, suivant le théorème des valeurs intermédiaires, la fonction $H(x)$ admet au moins un zéro dans l'intervalle $(0, 1)$ et au moins un zéro dans l'intervalle $(0, +\infty)$.

Le résultat suivant exprime le niveau de confiance défini sous forme d'un intervalle ou d'un ensemble d'intervalles disjoints.

RESULTAT 3

Si la fonction $H(x)$ possède exactement deux zéros, x_L et x_U , alors $0 < x_L < 1 < x_U$ et le niveau de confiance défini pour $\rho^2 = \sigma_1^2/\sigma_2^2$ est l'intervalle exprimé par la formule :

$$[c\hat{\rho}^2 x_L, c\hat{\rho}^2 x_U]$$

Par conséquent, l'IC pour $\rho = \sigma_1/\sigma_2$ est l'intervalle

$$[\hat{\rho}\sqrt{cx_L}, \hat{\rho}\sqrt{cx_U}]$$

Si en revanche la fonction $H(x)$ possède plus de deux zéros, la valeur de confiance définie pour $\rho^2 = \sigma_1^2/\sigma_2^2$ est fournie par un ensemble d'intervalles qui ne se chevauchent pas. Les bornes de chaque intervalle sont définies par les zéros consécutifs entre lesquels la fonction s'ouvre vers le haut.

Ce résultat est démontré dans l'Annexe C.

REMARQUE :

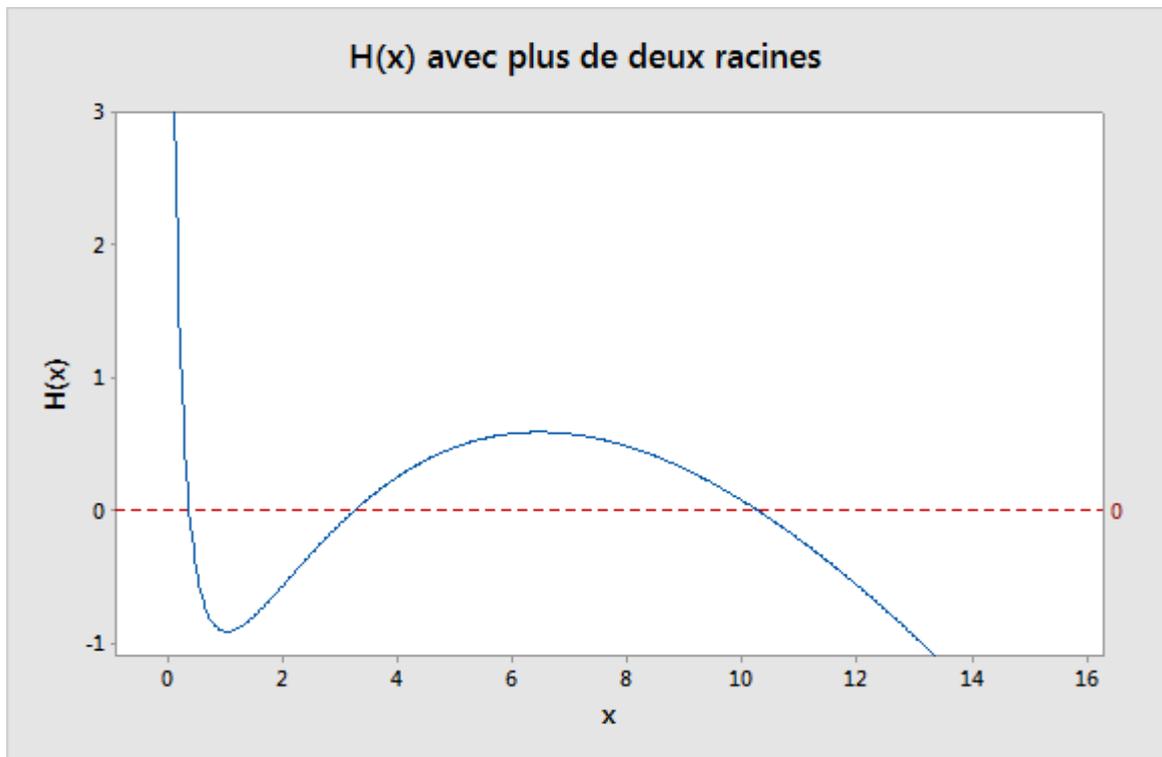
Bien qu'il soit mathématiquement possible que la fonction $H(x)$ admette plus de deux zéros, nous avons constaté que cela était uniquement le cas avec des plans très inhabituels et pratiquement dénués d'intérêt, où au moins un des échantillons est trop petit ou particulièrement non équilibré. Nous supposons que $H(x)$ possède deux ou quatre zéros.

L'exemple suivant utilise des données qui ont été créées de telle sorte que la fonction $H(x)$ possède obligatoirement plus de deux zéros. Les données sont résumées comme suit : $n_1 = 169$, $n_2 = 7$, $S_1 = 301,855$, $S_2 = 4606,170$, $\hat{\gamma}_1 = 1,877$, $\hat{\gamma}_2 = 6,761$, $c = 0,728$, $A = 30,381$, $B = 0,101$ et $K = 28,000$.

Pour $\alpha = 0,05$, la fonction $H(x)$ est exprimée comme suit :

$$H(x) = (\ln x)^2 - 1,960^2 \left(30,381 \frac{1,877 \times 28^2/169 + 6,761 \times (0,728x)^2/7}{(28,000 + 0,728x)^2} - 0,101 \right)$$

Dans ce cas, la fonction $H(x)$ possède quatre zéros. Le graphique de cette fonction est affiché ci-dessous. Vous remarquerez que le quatrième zéro n'est pas visible sur le graphique, car sa valeur est trop grande. Toutefois, nous savons que ce quatrième zéro existe, puisque $H(+\infty) = +\infty$.



La valeur numérique des quatre zéros est déterminée comme suit : $x_1 = 0,389$, $x_2 = 3,282$, $x_3 = 10,194$ et $x_4 = 39685,0$. Le rapport estimé des écarts types est de $\hat{\rho} = S_1/S_2 = 0,066$. Le niveau de confiance défini pour $\rho^2 = \sigma_1^2/\sigma_2^2$ peut être exprimé comme suit :

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] \cup [c \hat{\rho}^2 x_3, c \hat{\rho}^2 x_4] = [0,001, 0,010] \cup [0,032, 124,072]$$

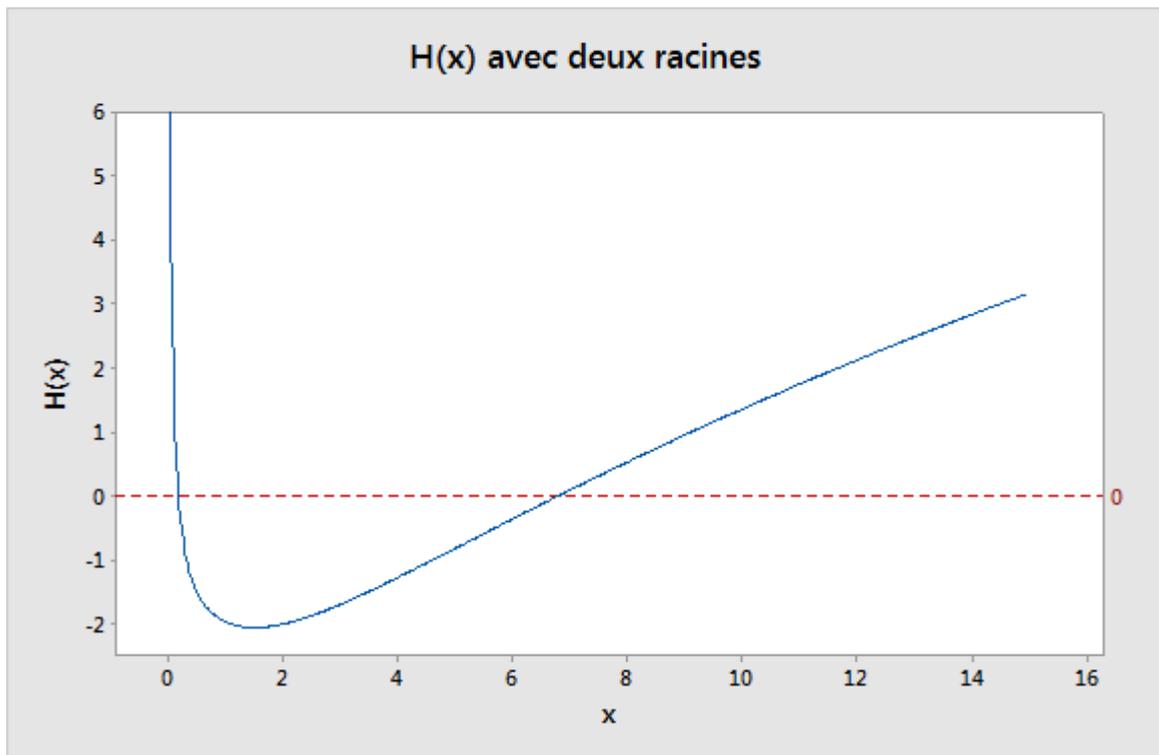
Le niveau de confiance défini pour le rapport des écarts types, ρ , est obtenu en calculant la racine carrée des bornes des intervalles.

Lorsque les échantillons ne sont pas trop petits ($n_i \geq 10$) et que l'écart entre leurs effectifs n'est pas trop élevé, la fonction $H(x)$ admet généralement deux zéros. Comme décrit dans le résultat 2, l'un des deux zéros est inférieur à 1, tandis que l'autre est supérieur à 1. L'exemple suivant s'appuie sur des données créées de façon aléatoire. Les données peuvent être résumées comme suit : $n_1 = 10$, $n_2 = 12$, $S_1 = 1,150$, $S_2 = 1,043$, $\hat{\gamma}_1 = 2,704$, $\hat{\gamma}_2 = 3,671$, $c = 1,041$, $A = 4,444$, $B = 0,146$ et $K = 0,818$.

Dans ce cas, pour $\alpha = 0,05$, la fonction $H(x)$ est exprimée comme suit :

$$H(x) = (\ln x)^2 - 1,960^2 \left(4,444 \frac{2,704 \times 0,818^2/10 + 3,671 \times (1,041x)^2/12}{(0,818 + 1,041x)^2} - 0,146 \right)$$

La fonction $H(x)$ possède deux zéros, comme indiqué ci-dessous :



La valeur numérique des deux zéros est calculée comme suit : $x_1 = 0,200$ et $x_2 = 6,824$. Le rapport des écarts types estimé est le suivant : $\hat{\rho} = S_1/S_2 = 1,102$. Le niveau de confiance défini pour $\rho^2 = \sigma_1^2/\sigma_2^2$ est l'intervalle exprimé comme suit :

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] = [0,253, 8,634]$$

L'IC pour le rapport des écarts types, ρ , est obtenu en calculant la racine carrée des bornes de l'intervalle indiqué ci-dessus.

Nous allons à présent décrire deux algorithmes permettant de déterminer les limites de confiance.

Le premier algorithme applique une procédure de recherche numérique des zéros afin de déterminer les points d'annulation (zéros) de la fonction $H(x)$. Le zéro correspondant à une limite de confiance inférieure pour le rapport des variances est compris dans l'intervalle $(0, 1)$. Si nous appelons ce zéro x_L , conformément au résultat 3, la limite de confiance inférieure pour le rapport des variances est fournie par la formule $c\hat{\rho}^2 x_L$ et celle pour le rapport des écarts types par la formule $\hat{\rho}\sqrt{cx_L}$. De la même façon, la limite de confiance supérieure pour le rapport des variances est de $c\hat{\rho}^2 x_U$ et celle pour le rapport des écarts types est de $\hat{\rho}\sqrt{cx_U}$, où $x_U > 1$ représente l'autre zéro de $H(x)$. Pour obtenir simplement la limite de confiance supérieure, il suffit de partir du principe que la limite inférieure de $1/\rho^2$ est la limite supérieure de ρ^2 . Dans un premier temps les rôles des deux échantillons sont inversés dans l'expression de la fonction $H(x)$, comme si nous calculions la limite de confiance pour le rapport $1/\rho^2 = \sigma_2^2/\sigma_1^2$. Ensuite, l'algorithme permettant d'obtenir la limite inférieure est appliqué à la nouvelle fonction $H(x)$. Enfin, la limite obtenue est inversée afin d'obtenir la limite de confiance supérieure souhaitée.

Une approche alternative consiste à calculer de façon récursive la limite de confiance inférieure pour le rapport des variances à l'aide de la relation de récurrence exprimée par la formule

$$\rho_0^2 = 1$$

$$\rho_{i+1}^2 = \exp \left[\ln(c \hat{\rho}^2) - z_{\alpha/2} \sqrt{\frac{\hat{\gamma}_P(\rho_i) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_i) - g_2}{n_2 - 1}} \right], i = 0, 1, 2, \dots$$

La limite de confiance inférieure pour le rapport des variances est de ρ_{j+1}^2 , avec $|\rho_{j+1}^2 - \rho_j^2| < \varepsilon$, où $j > 0$ et la valeur de ε est choisie pour être minimale (par exemple $\varepsilon = 10^{-6}$). Pour obtenir la limite de confiance supérieure, il suffit de remplacer $-z_{\alpha/2}$ par $+z_{\alpha/2}$ dans la formule ci-dessus.

Manifestement, les deux algorithmes de calcul des limites de confiance sont équivalents, car la procédure récursive est pour l'essentiel une procédure itérative permettant de résoudre l'équation $H(\rho^2/(c\hat{\rho}^2)) = 0$ pour ρ^2 . L'algorithme récursif étant plus facile à appliquer, il constitue donc une solution de secours utile lorsque vous ne disposez pas d'une procédure de recherche des zéros.

4. Etudes par simulation et résultats

Dans cet article, nous avons décrit une procédure visant à étendre le test d'égalité de deux variances de Layard afin de tester le rapport des variances. Nous appelons cette procédure Test de Layard étendu pour le rapport, ou ELTR (Extended Layard's Test for the Ratio). Dans cette section, nous étudions les propriétés des IC obtenus avec la procédure ELTR pour de petits échantillons. Nous suivons la procédure générale adoptée par Bonett (2006).

Nous comparons les IC obtenus avec la procédure ELTR à ceux calculés à partir des tests L_{50} (Pan, 1999) et W_{50} (test de Levene/Brown-Forsythe). Pour l'étude 1, nous incluons également les IC issus du test F classique, comme base de comparaison. En effet, il est connu que lorsque les données sont distribuées normalement, le test F classique obtient des résultats optimaux. Notez que les calculs utilisés pour déterminer les IC issus des tests W_{50} et L_{50} sont tirés de Pan (1999). Les calculs pour les IC issus du test F sont disponibles dans de nombreux manuels d'introduction aux statistiques. Ils sont également fournis par Bonett (2006).

Nous avons mené trois études par simulation, chacune incluant 100 000 répliques d'échantillons. Chaque réplique compte deux échantillons indépendants d'effectif faible ou modéré. Chaque échantillon a été créé à partir d'une loi parent ayant des propriétés connues (symétrie, asymétrie, queues lourdes/légères). L'erreur type associée à chaque simulation est d'environ 0,0009, 0,0007 et 0,0003 pour les niveaux de confiance nominaux de 90 %, 95 % et 99 %, respectivement.

Pour évaluer les résultats de chaque procédure, nous indiquons la probabilité de couverture obtenue et la largeur moyenne des intervalles simulés pour le rapport des variances. Certains des intervalles associés au test W_{50} présentent une largeur infinie (une possibilité exposée par Pan (1999)). Dans de tels cas, nous signalons à la fois la largeur moyenne des intervalles finis et le pourcentage d'intervalles ayant une largeur infinie. Toutes les simulations ont été effectuées à l'aide de la version 8 du logiciel Mathematica.

Etude 1 : comparaison des probabilités de couverture pour les données normales

Dans la première étude, nous créons des échantillons aléatoires de différents effectifs et obéissant à la loi normale. Les résultats sont présentés dans le tableau 2.

Tableau 2 : Comparaison des probabilités de couverture et de la largeur moyenne des intervalles.

$1 - \alpha$	n_1, n_2	Mesure	Procédure			
			F	ELTR	L_{50}	W_{50}
0,90	10, 10	Couverture	0,898	0,918	0,913	0,921
		Largeur	3,72	5,06	4,72	8,03 (0,01%)
	30, 10	Couverture	0,90	0,909	0,897	0,911
		Largeur	2,42	3,01	3,58	3,17
	25, 25	Couverture	0,902	0,907	0,914	0,916
		Largeur	1,61	1,73	1,85	1,938
	50, 50	Couverture	0,90	0,901	0,906	0,907
		Largeur	1,03	1,06	1,13	1,15
0,95	10, 10	Couverture	0,949	0,963	0,958	0,964
		Largeur	4,90	7,72	6,52	497,24 (0,20%)
	30, 10	Couverture	0,95	0,957	0,945	0,959
		Largeur	2,98	4,91	4,67	4,07
	25, 25	Couverture	0,951	0,955	0,958	0,961
		Largeur	1,99	2,24	2,31	2,49
	50, 50	Couverture	0,951	0,952	0,953	0,954
		Largeur	1,25	1,31	1,38	1,41

$1 - \alpha$	n_1, n_2	Mesure	Procédure			
			F	ELTR	L_{50}	W_{50}
0,99	10, 10	Couverture	0,989	0,993	0,992	0,994
		Largeur	8,29	17,76	12,52	$> 10^4$ (8.8%)
	30, 10	Couverture	0,99	0,992	0,986	0,994
		Largeur	4,26	15,76	8,26	6,77
	25, 25	Couverture	0,99	0,992	0,992	0,993
		Largeur	2,86	3,66	3,43	4,03
	50, 50	Couverture	0,99	0,991	0,991	0,991
		Largeur	1,71	1,89	1,92	2,02

Les lignes plus sombres indiquent les probabilités de couverture (Couverture) obtenues pour chaque procédure à chaque niveau de confiance ($1 - \alpha$) et pour chaque combinaison d'effectifs d'échantillons (n_1, n_2). La largeur moyenne des intervalles (Largeur) apparaît en dessous de chaque probabilité de couverture. Si les intervalles correspondant à une condition sont infinis, nous indiquons à la fois la moyenne des intervalles finis et le pourcentage d'intervalles infinis.

Comme prévu, les résultats indiquent que les IC associés à la procédure F sont les plus justes et les plus précis. Les probabilités de couverture obtenues avec la procédure F sont plus proches de la couverture cible que celles associées à d'autres procédures. De même, les largeurs moyennes d'intervalles associées à la procédure F sont inférieures à celles associées aux autres procédures. En outre, ce tableau révèle que les IC qui s'appuient sur les procédures ELTR et L_{50} sont presque aussi exacts et aussi précis que ceux fondés sur la procédure F.

Les intervalles qui s'appuient sur le test W_{50} offrent eux aussi un bon niveau d'exactitude. Toutefois, selon l'effectif des échantillons, les intervalles W_{50} peuvent être très larges, voire présenter une largeur infinie. Notez que lorsque les deux échantillons ne présentent que 10 observations, au moins 0,01 % des intervalles générés avec la procédure W_{50} ont une largeur infinie. En outre, lorsque la couverture cible augmente, le pourcentage d'intervalles infinis augmente aussi. Dans la plupart des conditions, les largeurs moyennes d'intervalles obtenues avec les procédures ELTR et L_{50} sont inférieures à celles obtenues avec la méthode W_{50} .

Etude 2 : comparaison des probabilités de couverture pour les données non normales

L'objectif de la deuxième étude est d'évaluer et de comparer les résultats des procédures ELTR, L_{50} et W_{50} avec des lois parent non normales. Nous incluons également une loi normale contaminée afin d'évaluer l'impact des valeurs aberrantes sur les performances des

procédures. Nous écrivons cette distribution contaminée sous la forme $CN(0,1, 3)$ pour indiquer que même si 90 % des observations sont créées à partir de la loi normale standard, les 10 % restants sont issus d'une population normale avec une moyenne de 0 et un écart type de 3. Les résultats sont présentés dans le tableau 3.

Tableau 3 : Comparaison des probabilités de couverture et de la largeur moyenne des intervalles dans des modèles non normaux. Le niveau de confiance nominal est de $1 - \alpha = 0,95$.

Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}
Uniforme				$\chi^2(5)$			
[1,8]				[5,4]			
10, 10	0,971	0,971	0,966	10, 10	0,956	0,938	0,956
	5,27	4,87	42,08 (0.1%)		11,61	8,78	> 10 ⁴ (2.6%)
10, 30	0,964	0,961	0,957	10, 30	0,959	0,923	0,956
	2,51	2,4	2,89		6,25	4,14	190,645 (0.3%)
25, 25	0,967	0,972	0,968	25, 25	0,956	0,944	0,954
	1,43	1,79	1,88		3,66	2,92	3,26
50, 50	0,959	0,962	0,959	50, 50	0,959	0,946	0,952
	0,83	1,06	1,08		2,07	1,7	1,77
Bêta(3, 3)				Exponentielle			
[2,5]				[9]			
10, 10	0,968	0,966	0,966	10, 10	0,947	0,916	0,95
	6,26	5,59	254,62 (0.1%)		20,99	14,47	> 10 ⁴ (9.1%)
10, 30	0,96	0,954	0,96	10, 30	0,954	0,896	0,953
	3,14	2,76	3,71		10,46	6,19	> 10 ⁴ (4.1%)
25, 25	0,959	0,966	0,965	25, 25	0,956	0,931	0,951
	1,81	2,06	2,18		6,09	4,13	5,48 (0.008%)
50, 50	0,957	0,959	0,958	50, 50	0,962	0,942	0,952
	1,06	1,23	1,26		3,18	2,24	2,38

Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}
Laplace				$\chi^2(1)$			
[6]				[15]			
10, 10	0,946	0,935	0,961	10, 10	0,928	0,889	0,947
	13,47	10,45	$> 10^4$ (3.0%)		55,09	37,4	$> 10^5$ (25.1%)
10, 30	0,947	0,919	0,957	10, 30	0,943	0,882	0,956
	6,78	4,82	$> 10^4$ (0.4%)		18,71	11,14	$> 10^6$ (25.7%)
25, 25	0,945	0,94	0,952	25, 25	0,952	0,925	0,954
	4,00	3,372	3,86		10,97	6,84	$> 10^4$ (0.4%)
50, 50	0,952	0,949	0,955	50, 50	0,958	0,936	0,951
	2,19	1,91	1,99		5,08	3,31	3,75 (0.001%)
t(5)				Log-normale			
[9]				[113,9]			
10, 10	0,957	0,946	0,965	10, 10	0,923	0,876	0,955
	11,07	8,81	$> 10^3$ (2.0%)		59,22	46,15	$> 10^5$ (23.0%)
10, 30	0,957	0,93	0,959	10, 30	0,949	0,866	0,958
	6,06	4,24	$> 10^3$ (0.7%)		29,13	17,67	$> 10^6$ (31.6%)
25, 25	0,954	0,948	0,96	25, 25	0,947	0,917	0,965
	3,54	2,93	4,86 (0.01%)		16,21	8,73	$> 10^4$ (2.4%)
50, 50	0,954	0,947	0,954	50, 50	0,955	0,928	0,96
	2,10	1,71	1,77 (0.003%)		8,62	4,11	164,38 (0.2%)

Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Loi de distribution [γ] n_1, n_2	ELTR	L_{50}	W_{50}
Demi-normale				CN(0,1, 3)			
[3,9]				[8,3]			
10, 10	0,956	0,942	0,954	10, 10	0,977	0,965	0,979
	10,41	7,89	$> 10^4$ (1.5%)		12,64	9,52	$> 10^4$ (4.9%)
10, 30	0,959	0,93	0,954	10, 30	0,981	0,952	0,979
	5,18	3,64	13,00 (0,02%)		7,82	4,71	944,68 (1,1%)
25, 25	0,959	0,952	0,959	25, 25	0,982	0,972	0,981
	3,01	2,62	2,88		4,63	3,22	3,71
50, 50	0,96	0,951	0,954	50, 50	0,983	0,972	0,978
	1,69	1,54	1,59		2,64	1,83	1,91

Les lignes plus sombres indiquent les probabilités de couverture obtenues pour chaque procédure, loi parent et combinaison d'effectifs d'échantillons. La largeur moyenne des intervalles apparaît en dessous de chaque probabilité de couverture. Si les intervalles correspondant à une condition sont infinis, nous indiquons à la fois la moyenne des intervalles finis et le pourcentage d'intervalles infinis. L'aplatissement (γ) de chaque loi parent est affiché entre crochets.

Pour les lois de distribution symétriques à queues légères, les résultats indiquent que les trois méthodes présentent des probabilités de couverture prudentes. Toutefois, les intervalles ELTR et L_{50} sont plus précis pour les petits échantillons que les intervalles W_{50} . Par exemple, lorsque les échantillons sont créés à partir d'une loi bêta avec les paramètres (3, 3), les probabilités de couverture obtenues pour les intervalles ELTR et L_{50} sont au moins aussi précises que celles des intervalles W_{50} . En revanche, les intervalles W_{50} sont invariablement plus larges.

Les intervalles ELTR et W_{50} sont aussi relativement prudents pour les lois symétriques à queues lourdes, tandis que les intervalles L_{50} sont libéraux. Les intervalles L_{50} sont encore plus libéraux dans les plans non équilibrés. Par exemple, lorsque des échantillons ayant des effectifs de 10 et 30 sont créés à l'aide de la loi de Laplace, la probabilité de couverture obtenue pour les intervalles L_{50} est de 0,919. Pour des échantillons de même effectif créés à partir d'une loi t à 5 degrés de liberté, la probabilité de couverture obtenue pour les intervalles L_{50} est de 0,930.

Les intervalles L_{50} sont également très libéraux pour de petits échantillons obéissant à des lois très asymétriques à queues lourdes. Par exemple, lorsque les échantillons sont issus de populations obéissant à une loi log-normale, la couverture obtenue peut n'être que de 0,866. Pour ces lois, la méthode W_{50} est la moins libérale des trois. Toutefois, il existe un nombre

trop élevé d'intervalles W_{50} qui présentent une largeur infinie. Par exemple, lorsque des échantillons sont créés à partir de la loi du Khi deux à 1 degré de liberté ($\chi^2(1)$), plus de 25 % des intervalles W_{50} peuvent avoir une largeur infinie. Les intervalles ELTR sont un peu moins précis, mais considérablement plus étroits et donc plus informatifs que les intervalles W_{50} .

Enfin, nous constatons que les valeurs aberrantes ont une incidence négative sur les trois procédures. La méthode L_{50} est la moins sensible à ces valeurs aberrantes. Cela était prévisible dans la mesure où la méthode L_{50} a été adaptée pour réduire l'effet des valeurs aberrantes sur le test W_{50} (Pan, 1999). Lorsque les échantillons sont créés à partir de la loi normale contaminée, $CN(0,1, 3)$, la valeur de probabilité de couverture minimale obtenue pour les procédures ELTR et W_{50} est de 0,977. Les résultats de simulations supplémentaires (non indiqués) indiquent que la qualité de ces intervalles ne s'améliore que lentement lorsque les effectifs d'échantillons augmentent.

Etude 3 : sensibilité à l'hypothèse d'égalité des aplatissements

Cette dernière étude examine la sensibilité de la procédure ELTR à l'hypothèse d'égalité des aplatissements sous laquelle elle a été formulée. Nous examinons les résultats de la procédure ELTR lorsque les aplatissements des populations parent sont inégaux, c'est-à-dire lorsque $\gamma_1 \neq \gamma_2$. Nous incluons également les procédures L_{50} et W_{50} , car elles s'appuient sur l'hypothèse selon laquelle les populations sont similaires. Cette hypothèse de similarité est mise à mal lorsque les aplatissements des populations parent sont inégaux. Les résultats sont présentés dans le tableau 4.

Tableau 4 : Sensibilité de la procédure ELTR à l'hypothèse d'égalité des aplatissements. Le niveau de confiance nominal est de $1 - \alpha = 0,95$.

Loi 1, Loi 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	Loi 1, Loi 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
Bêta(3, 3), Normale [2,5, 3]				Normale, CN(0,9, 3) [3, 8,3]			
10, 10	0,964	0,961	0,964	10, 10	0,955	0,948	0,951
	0,27	0,23	204,50 (0,20%)		6,88	5,16	> 10 ⁴ (4,89%)
30, 10	0,946	0,939	0,946	30, 10	0,941	0,91	0,942
	0,16	0,17	0,15		5,26	3,77	3,20
10, 30	0,966	0,956	0,967	10, 30	0,961	0,95	0,958
	0,14	0,11	0,17		4,26	2,40	630,42 (1,10%)
50, 50	0,951	0,95	0,949	50, 50	0,936	0,91	0,907

Loi 1, Loi 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	Loi 1, Loi 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
	0,04	0,05	0,05		1,27	1,11	1,19
Normale, Laplace [3, 6]				Demi- normale, $\chi^2(5)$ [3,9, 5,4]			
10, 10	0,941	0,935	0,947	10, 10	0,956	0,94	0,954
	6,67	5,17	$> 10^6$ (2.90%)		0,42	0,32	304,41 (2.60%)
30, 10	0,912	0,888	0,914	30, 10	0,954	0,918	0,949
	5,06	3,85	3,21		0,33	0,22	0,20
10, 30	0,963	0,943	0,955	10, 30	0,962	0,934	0,958
	3,33	2,25	$> 10^3$ (0.40%)		0,23	0,15	3,28 (0.30%)
50, 50	0,935	0,894	0,889	50, 50	0,955	0,941	0,945
	0,98	1,04	1,12		0,07	0,06	0,07
Normale, Demi-normale [3, 3,9]				$\chi^2(5),$ Exponentielle [5,4, 9]			
10, 10	0,956	0,948	0,957	10, 10	0,938	0,914	0,94
	28,16	20,65	$> 10^4$ (1.50%)		211,17	137,88	$> 10^6$ (9.10%)
30, 10	0,946	0,924	0,947	30, 10	0,928	0,875	0,929
	20,59	14,83	12,78		194,70	93,02	83,02
10, 30	0,961	0,946	0,962	10, 30	0,968	0,93	0,954
	14,06	9,37	49,11 (0.02%)		102,35	55,29	$> 10^5$ (3.90%)
50, 50	0,953	0,95	0,952	50, 50	0,95	0,92	0,923
	4,32	4,16	4,33		29,64	23,37	25,54

Les lignes plus sombres indiquent les probabilités de couverture obtenues pour chaque procédure, combinaison de lois parent (Loi 1, Loi 2) et combinaison d'effectifs d'échantillons. La largeur moyenne des intervalles apparaît en dessous de chaque probabilité de couverture. Si les intervalles correspondant à une condition sont infinis, nous indiquons à la fois la moyenne des intervalles finis et le pourcentage d'intervalles infinis. L'aplatissement (γ_1, γ_2) de chaque loi parent est affiché entre crochets.

En général, lorsque les échantillons sont suffisamment grands, l'inégalité des aplatissements ne semble pas avoir d'incidence négative sur les résultats de la procédure ELTR. Toutefois, lorsque les plans sont non équilibrés et que le plus petit échantillon obéit à une loi à queues lourdes, les probabilités de couverture obtenues sont libérales. Les probabilités de couverture sont meilleures lorsque l'échantillon le plus grand est créé à partir d'une loi à queues lourdes.

Lorsque les échantillons sont suffisamment grands, les intervalles L_{50} et W_{50} semblent généralement robustes aux différences de distribution résultant d'aplatissements inégaux. Toutefois, notez que lorsque les échantillons sont créés à partir d'une loi normale et d'une loi de Laplace, ou d'une loi normale et d'une loi normale contaminée, les probabilités de couverture des intervalles L_{50} et W_{50} ne sont pas stables, même pour des échantillons aussi grands que 50.

Les intervalles L_{50} sont généralement plus libéraux que les intervalles ELTR et W_{50} . Dans trois cas, les probabilités de couverture obtenues pour les intervalles L_{50} sont inférieures à 0,90. En revanche, seule une des probabilités de couverture obtenues pour les intervalles W_{50} est inférieure à 0,90. La probabilité de couverture la plus faible obtenue pour les intervalles ELTR est égale à 0,912.

L'étude précédente (tableau 3) indique que les trois procédures produisent des intervalles beaucoup plus prudents lorsque les deux échantillons sont créés à partir de la loi normale contaminée, $CN(0,1, 3)$. Cette étude montre que les trois procédures donnent de meilleurs résultats lorsqu'un seul échantillon est créé à partir de la loi $CN(0,1, 3)$. Toutefois, notez que les performances des intervalles L_{50} et W_{50} semblent diminuer considérablement lorsque l'effectif de l'échantillon atteint 50.

5. Exemple

Dans cette section, nous appliquons les quatre procédures (F, ELTR, L_{50} et W_{50}) à un ensemble de données fournies par Pan (1999). Ott (1993, page 352) décrit les données comme suit :

Dans une mine de fer, un chimiste soupçonne que la variance dans la quantité (le poids, en onces) d'oxyde de fer par livre de minerai tend à augmenter à mesure que la quantité moyenne d'oxyde de fer par livre augmente. Pour tester cette théorie, dix échantillons d'1 livre de minerai de fer sont chacun prélevés dans deux emplacements. Le premier, l'emplacement 1, présente une teneur moyenne en oxyde de fer beaucoup plus élevée que le second, l'emplacement 2. Les quantités d'oxyde de fer contenu dans les échantillons de minerai sont présentées ci-dessous :

Emplacement 1	8,1	7,4	9,3	7,5	7,1	8,7	9,1	7,9	8,4	8,8
Emplacement 2	3,9	4,4	4,7	3,6	4,1	3,9	4,6	3,5	4,0	4,2

Les IC à 95 % calculés pour $\sigma_2/\sigma_1 = 1/\rho$ à l'aide des quatre méthodes sont présentés dans le tableau suivant :

Procédure	IC à 95 %
F	(0,262, 1,055)

Procédure	IC à 95 %
ELTR	(0,277, 0,924)
L_{50} (Pan)	(0,295, 0,938)
W_{50} (Levene/Brown-Forsythe)	(0,237, 0,908)

6. Conclusion

Nos simulations montrent que, de manière générale, les IC obtenus avec la procédure ELTR présentent un niveau d'exactitude équivalent à ceux calculés à partir des tests L_{50} et W_{50} . En revanche, pour la plupart des lois, les intervalles ELTR et L_{50} sont plus précis que les intervalles W_{50} . Les intervalles W_{50} ont tendance offrir une plus grande exactitude que les intervalles ELTR et L_{50} avec de petits échantillons créés à partir de lois très asymétrique à queues lourdes. Toutefois, cet avantage est généralement contrebalancé par une perte de précision considérable. Les intervalles W_{50} obtenus sont généralement trop larges et susceptibles de présenter une étendue infinie.

Les intervalles L_{50} améliorent bien la précision des intervalles W_{50} , ce qui était leur objectif. Toutefois, pour les populations asymétriques, les intervalles L_{50} sont si étroits qu'ils produisent des probabilités de couverture trop libérales. Par comparaison, les intervalles ELTR sont généralement plus stables. Les intervalles ELTR ne sont habituellement ni trop larges, ni trop étroits, ce qui permet d'obtenir des probabilités de couverture ni trop prudentes ni trop libérales. Par conséquent, la procédure ELTR semble être la plus adaptée pour la plupart des applications pratiques.

Les intervalles ELTR requièrent certes un peu plus de calculs que les intervalles fondés sur les tests L_{50} ou W_{50} . Toutefois, de manière générale, le gain de précision (par rapport aux intervalles W_{50}) et d'exactitude (par rapport aux intervalles L_{50}) compense largement l'effort de calcul supplémentaire. La procédure ELTR a été appliquée à l'analyse de la variance à deux échantillons dans la version 17 de Minitab Statistical Software, où elle est appelée procédure de Bonett.

Pour une recherche future, il serait intéressant d'étudier les propriétés du test de Layard pour de petits échantillons, dans des plans à échantillons multiples où l'estimateur d'aplatissement regroupé de Layard serait remplacé par la version plus robuste de Bonett, exprimée comme suit :

$$\hat{\gamma}_A = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^2 \right]^2} \sum_{i=1}^k n_i$$

où m_i est la moyenne tronquée de l'échantillon i , avec la proportion de troncature $1/[2(n_i - 4)^{1/2}]$, et $i = 1, \dots, k$.

En outre, il pourrait être judicieux d'utiliser l'approximation de la variance asymptotique de la variance de l'échantillon transformé par logarithme.

Enfin, nous constatons que les intervalles proposés par Bonett (2006), bien qu'ils ne puissent être utilisés comme de véritables IC, n'en sont pas moins remarquablement exacts et précis

pour la plupart des lois lorsqu'ils sont interprétés comme des zones d'acceptation pour le test d'égalité de deux variances. Ces zones d'acceptation sont idéales pour comparer plusieurs variances de façon graphique. Hochberg, Weiss et Hart (1982) ont proposé une procédure similaire pour tester l'égalité des moyennes. Cette procédure a été appliquée à l'analyse du test d'égalité des variances dans la version 17 de Minitab Statistical Software, où elle est appelée procédure de comparaisons multiples.

7. Annexes

Annexe A : démonstration du résultat 1

Partant des notations et des hypothèses de la section 2, soit $X_j = \rho Y_{2j}$ pour une valeur $\rho = \sigma_1/\sigma_2$ donnée. Alors

$$\text{Var}(X_j) = \rho^2 \text{Var}(Y_{2j}) = \rho^2 \sigma_2^2 = \sigma_1^2 = \text{Var}(Y_{1j})$$

et

$$E\left(\frac{X_j - \mu_{X_j}}{\sigma_{X_j}^2}\right)^4 = \rho^4 E(Y_{2j} - \mu_2) / (\rho^4 \sigma_2^4) = E(Y_{2j} - \mu_2) / \sigma_2^4 = \gamma$$

Etant donné que par hypothèse $E(Y_{1j} - \mu_1) / \sigma_1^4 = \gamma$, les populations parent des deux échantillons Y_{1j} et $X_j = \rho Y_{2j}$ présentent la même variance σ_1^2 et le même aplatissement γ : D'après Layard (1973), nous pouvons exprimer un estimateur cohérent de l'aplatissement regroupé de γ pour deux échantillons Y_{1j} et X_j comme suit :

$$\begin{aligned} \hat{\gamma}' &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \sum_{j=1}^{n_2} (X_j - \bar{X})^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_X^2]^2} \\ &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = \hat{\gamma}_P(\rho) \end{aligned}$$

conformément à nos résultats.

Annexe B : démonstration du résultat 2

Nous avons déjà établi qu'un niveau de confiance approximatif de $(1 - \alpha)100$ pour cent défini pour $\rho = \sigma_1/\sigma_2$ avec T_2 peut être exprimé comme suit :

$$\left\{ \rho \in (0, \infty) : (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \leq 0 \right\}$$

L'estimateur d'aplatissement regroupé peut être exprimé en fonction des estimateurs d'aplatissement des échantillons individuels, exprimés comme suit :

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

Plus spécifiquement, si $t = \rho/\hat{\rho}$, alors

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = (n_1 + n_2) \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2}$$

où $K = (n_1 - 1)/(n_2 - 1)$.

Par conséquent, le carré du terme d'erreur type peut être exprimé comme suit :

$$\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2} - B$$

où

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{k_1}{n_1 - 1} + \frac{k_2}{n_2 - 1}$$

Par conséquent, pour $r^2 = \rho^2/(c\hat{\rho}^2)$, nous pouvons facilement observer que :

$$\begin{aligned} (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \\ = (\ln r^2)^2 - z_{\alpha/2}^2 \left(A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 c^2 r^4/n_2}{(K + c r^2)^2} - B \right) \end{aligned}$$

Ainsi, un niveau de confiance approximatif de $(1 - \alpha)100$ pour cent défini pour $\rho = \sigma_1/\sigma_2$ avec T_2 peut être exprimé comme suit :

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

ou, de façon équivalente, le niveau de confiance défini pour $\rho^2 = \sigma_1^2/\sigma_2^2$ peut être exprimé comme suit :

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

où

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 es^2(cx), x > 0$$

et

$$es^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

Annexe C : démonstration du résultat 3

Il est facile de vérifier que la fonction $H(x)$ est continue sur la droite des réels positifs, avec $H(0) = H(+\infty) = +\infty$ et $H(1) < 0$. Par conséquent, suivant le théorème des valeurs intermédiaires, la fonction $H(x)$ admet au moins un zéro dans l'intervalle $(0, 1)$ et au moins un zéro dans l'intervalle $(0, +\infty)$. Par conséquent, si la fonction $H(x)$ possède exactement deux zéros, l'un est inférieur à 1 et l'autre est supérieur à 1. Comme cette fonction s'ouvre vers le haut, la relation d'inégalité $H(r) \leq 0$ est satisfaite lorsque r est compris entre les points d'annulation (zéros). Ces zéros définissent les bornes des IC pour $\rho^2/(c\hat{\rho}^2)$. Ainsi, si nous appelons $x_L < 1 < x_U$ les deux zéros, d'après le résultat 2, la limite de confiance inférieure pour le rapport des variances, ρ^2 , est fournie par la formule $c\hat{\rho}^2 x_L$ et celle pour le rapport des écarts types par la formule $\hat{\rho}\sqrt{cx_L}$. De la même façon, la limite de confiance supérieure pour le rapport des variances est de $c\hat{\rho}^2 x_U$ et celle pour le rapport des écarts types est de $\hat{\rho}\sqrt{cx_U}$.

En revanche, si la fonction $H(x)$ possède plus de deux racines, la relation d'inégalité $H(r) \leq 0$ est satisfaite lorsque r est compris entre des racines consécutives à l'intérieur desquelles la

fonction s'ouvre vers le haut. Par conséquent, le niveau de confiance défini est un ensemble d'intervalles qui ne se chevauchent pas.

8. Références

Balakrishnan, N. et Ma, C. W. (1990), A Comparative Study of Various Tests for the Equality of Two Population Variances, *Journal of Statistical Computation and Simulation*, 35, 41–89.

Bonett D. G. (2006), Robust Confidence Interval for a Ratio of Standard Deviations, *Applied Psychological Measurements*, 30, 432–439.

Boos, D. D. et Brownie, C. (1989), Bootstrap Methods for Testing Homogeneity of Variances, *Technometrics*, 31, 69–82.

Brown, M. B. et Forsythe, A. B. (1974), Robust Tests for the Equality of Variances, *Journal of the American Statistical Association*, 69, 364–367.

Conover, W. J., Johnson, M. E. et Johnson, M. M. (1981), A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data, *Technometrics*, 23, 351–361.

Fligner, M. A. et Killeen, T. J. (1976), Distribution-Free Two-Sample Tests for Scale, *Journal of the American Statistical Association*, 71, 210–213.

Hall, P. et Padmanabhan, A. R. (1997), Adaptive Inference for the Two-Sample Scale Problem, *Technometrics*, 39, 412–422.

Hochberg, Y., Weiss, G. et Hart S. (1982), On Graphical Procedures for Multiple Comparisons, *Journal of the American Statistical Association*, 77, 767–772.

Layard, M. W. J. (1973), Robust Large-Sample Tests for Homogeneity of Variances, *Journal of the American Statistical Association*, 68, 195–198.

Levene, H. (1960), "Robust Tests for Equality of Variances," in I. Olkin, éd., *Contributions to Probability and Statistics*, Palo Alto, CA : Stanford University Press, 278–292.

Lim, T.-S. et Loh, W.-Y. (1996), A Comparison of Tests of Equality of Variances, *Computational Statistics and Data Analysis*, 22, 287–301.

Ott, L. (1993), *An Introduction to Statistical Methods and Data Analysis*, Belmont, CA : Duxbury Press.

Pan, G. (1999), On a Levene Type Test for Equality of Two Variances, *Journal of Statistical Computation and Simulation*, 63, 59–71.

Shoemaker, L. H. (2003), Fixing the F Test for Equal Variances, *The American Statistician*, 57, 105–114.

Wolfram, S. (1999), *The Mathematica Book*, 4ème éd. Wolfram Media/Cambridge University Press.

© 2020 Minitab, LLC. All rights reserved. Minitab®, Minitab Workspace™, Companion by Minitab®, Salford Predictive Modeler®, SPM®, and the Minitab® logo are all registered trademarks of Minitab, LLC, in the United States and other countries. Additional trademarks of Minitab, LLC can be found at www.minitab.com. All other marks referenced remain the property of their respective owners.