

Méthode de comparaisons multiples

PROCEDURE GRAPHIQUE DE COMPARAISONS MULTIPLES POUR PLUSIEURS ECARTS TYPES,

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

ABSTRACT

L'article propose une nouvelle procédure graphique de comparaisons multiples de k écarts types. Utilisée comme test de l'homogénéité des variances, cette nouvelle procédure présente des propriétés semblables à celles de la version de Brown et Forsythe (1974) du test de Levene (1960), le test W_{50} , pour les erreurs de 1ère et 2ème espèces. Toutefois, l'affichage graphique associé au test de comparaisons multiples offre un outil visuel pratique pour le criblage d'échantillons présentant différents écarts types.

Termes d'index : homogénéité des variances, test de Levene, test de Brown-Forsythe, test de Layard, comparaisons multiples

1. Introduction

La version modifiée par Brown et Forsythe (1974) du test de Levene (1960), aussi appelée test W_{50} , est probablement l'une des procédures les plus couramment utilisées pour tester l'homogénéité (l'égalité) des variances. Ce test W_{50} est notamment apprécié pour sa robustesse et parce que, asymptotiquement, il ne dépend pas de lois de distribution. Par comparaison avec d'autres tests d'homogénéité des variances, le test W_{50} est facile à calculer. (Pour une comparaison de ces tests, voir Conove et al. (1981).) En outre, le test W_{50} est facilement accessible car il est disponible dans de nombreux logiciels statistiques tels que SAS, Minitab, R et JMP.

Toutefois, pour certaines lois, la puissance du test W_{50} peut être très faible, en particulier dans de petits échantillons. Par exemple, Pan (1999) observe que pour certaines lois, notamment la loi normale, le test W_{50} peut ne pas être suffisamment puissant pour détecter des différences entre deux écarts types, quelle que soit l'importance de ces différences.

L'analyse de Pan ne nous permet pas de déterminer si la même limitation s'applique aux plans à échantillons multiples. Nous pourrions penser que cette limitation ne s'applique pas aux plans à plus de deux échantillons, car ces derniers incluent a priori plus de données que les plans à deux échantillons. Nous savons que les propriétés du test W_{50} sont satisfaisantes pour les grands échantillons (Miller, 1968 ; Brown et Forsythe, 1974 ; Conover et al., 1981).

Lorsqu'un test W_{50} est significatif, il est devenu courant d'effectuer une procédure de comparaisons deux à deux simultanée en utilisant une correction de multiplicité de Bonferroni. Toutefois, comme le souligne Pan (1999), cette méthode à toutes les chances d'échouer ou de fournir des résultats erronés, en raison de la faible puissance du test W_{50} dans les plans à deux échantillons. La correction de Bonferroni, de par son caractère prudent, aggrave le problème, en particulier lorsque le nombre de comparaisons deux à deux est élevé. Cela dit, de nombreuses procédures de comparaisons multiples permettent de comparer efficacement des moyennes en effectuant une ANOVA à un facteur contrôlé. Pour des exemples, voir Tukey (1953), Hochberg et al. (1982) et Stoline (1981). Une analyse a posteriori analogue permettant de comparer des variances d'échantillons serait utile.

Dans cet article, nous proposons une méthode graphique de comparaison des variances (ou des écarts types) de plusieurs échantillons. L'analyse s'appuie sur des "intervalles d'incertitude" pour les variances, semblables aux intervalles d'incertitude décrits par Hochberg et al. (1982) pour les moyennes. Dans un premier temps, la procédure de comparaisons multiples deux à deux s'appuie sur la version modifiée par Bonett (2006) du test d'égalité des variances de Layard (1973) pour les plans à deux échantillons. La correction de la multiplicité utilisée dans les comparaisons deux à deux s'appuie sur une généralisation pour grands échantillons de la méthode de Tukey-Kramer (Tukey, 1953 ; Kramer, 1956), proposée par Nakayama (2009). Les intervalles d'incertitude, aussi appelés "intervalles de comparaisons multiples" ou "intervalles CM", sont dérivés à partir de la procédure de comparaisons deux à deux à l'aide de la procédure de meilleure approximation décrite par Hochberg et al. (1982). Le test de comparaisons multiples obtenu ne rejette l'hypothèse nulle que si, et seulement si, au moins deux intervalles de comparaisons multiples ne se chevauchent pas. Les intervalles de comparaisons multiples qui ne se chevauchent pas permettent d'identifier les échantillons dont les variances (ou les écarts types) sont significativement différentes.

Nous effectuons des simulations pour évaluer les propriétés du test de comparaisons multiples pour de petits échantillons. Comme base de comparaison, nous incluons également le test W_{50} dans nos études de simulation.

2. Procédure graphique de comparaisons multiples

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 $E(Y_{il}) = \mu_i$ et une variance $\text{Var}(Y_{il}) = \sigma_i^2 > 0$. Nous supposons en outre que les échantillons proviennent de populations présentant un aplatissement commun $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$.

Soit également \bar{Y}_i et S_i la moyenne et l'écart type de l'échantillon i , respectivement. Soit m_i la moyenne tronquée de l'échantillon i avec la proportion de troncature $1/[2\sqrt{n_i - 4}]$ et $\hat{\gamma}_{ij}$ l'estimateur d'aplatissement regroupé des échantillons (i, j) , exprimé de la façon suivante

$$\hat{\gamma}_{ij} = (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{\left[\sum_{l=1}^{n_i} (Y_{il} - \bar{Y}_i)^2 + \sum_{l=1}^{n_j} (Y_{jl} - \bar{Y}_j)^2 \right]^2}$$

$$= (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{\left[(n_i - 1)S_i^2 + (n_j - 1)S_j^2 \right]^2}$$

Notez que $\hat{\gamma}_{ij}$ est asymptotiquement équivalent à l'estimateur d'aplatissement regroupé de Layard (1973), où la moyenne de l'échantillon \bar{Y}_i a été remplacée par la moyenne tronquée m_i . Ainsi, $\hat{\gamma}_{ij}$ demeure un estimateur cohérent de l'aplatissement regroupé inconnu γ , tant que les variances des populations sont égales. Bonnett (2006) propose d'utiliser cet estimateur plutôt que celui de Layard pour améliorer les performances du test de Layard pour de petits échantillons dans des problèmes à deux échantillons. Tout au long de cette étude, nous désignons la version modifiée par Bonnett (2006) du test de Layard comme le test de Bonnett.

Supposons que nous devons comparer plus de deux échantillons ou groupes indépendants ($k > 2$). La procédure graphique de comparaisons multiples que nous proposons est inspirée des comparaisons multiples deux à deux fondées sur le test de Bonnett. Il est également possible de fonder les comparaisons deux à deux sur le test W_{50} . Toutefois, dans les plans à deux échantillons, la puissance du test W_{50} pose problème avec certaines lois, notamment la loi normale (Pan, 1999). En outre, Banga et Fox (2013) démontrent que les intervalles de confiance du test de Bonnett pour le rapport des variances sont généralement supérieurs à ceux du test W_{50} .

Pour toute paire d'échantillons (i, j) , un test de Bonnett bilatéral avec un seuil de signification de α' rejette l'hypothèse nulle d'égalité des variances si, et seulement si

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > z_{\alpha'/2} \sqrt{\frac{\hat{\gamma}_{ij} - k_i}{n_i - 1} + \frac{\hat{\gamma}_{ij} - k_j}{n_j - 1}}$$

où $z_{\alpha'/2}$ est le $\alpha'/2 \times 100$ ème point de percentile supérieur de la loi normale standard.

$$k_i = \frac{n_i - 3}{n_i}, k_j = \frac{n_j - 3}{n_j}, c_i = \frac{n_i}{n_i - z_{\alpha/2}}, c_j = \frac{n_j}{n_j - z_{\alpha/2}}$$

Puisque nous nous trouvons dans le cas de comparaisons multiples deux à deux, $k(k - 1)/2$ comparaisons exactement, il est nécessaire d'effectuer un ajustement de la multiplicité. Par exemple, lorsqu'un seuil de signification cible α global ou par famille est fourni, il est courant d'appliquer une procédure appelée correction de Bonferroni, consistant à choisir le seuil de signification de chacune des $k(k - 1)/2$ comparaisons deux à deux. La correction de Bonferroni est cependant connue pour produire des comparaisons deux à deux de plus en plus prudentes à mesure que le nombre d'échantillons à comparer augmente. Nakayama (2009) propose une approche plus satisfaisante, fondée sur une approximation pour grands échantillons de la méthode de Tukey-Kramer (Tukey, 1953 ; Kramer, 1956). Plus spécifiquement, le test de comparaisons multiples deux à deux global est significatif si, et seulement si, la condition suivante est satisfaite pour une paire d'échantillons (i, j) :

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > \frac{q_{k,\alpha}}{\sqrt{2}} \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

où $q_{\alpha,k}$ est le point α le plus élevé de l'étendue de k variables aléatoires, indépendantes et distribuées de façon identique selon la loi normale standard. C'est-à-dire, si $q_{\alpha,k}$ satisfait

$$\Pr\left(\max_{1 \leq i < j \leq k} |Z_i - Z_j| \leq q_{\alpha,k}\right) = 1 - \alpha$$

où Z_1, \dots, Z_k sont des variables aléatoires indépendantes, distribuées de façon identique selon la loi normale standard. Barnard (1978) fournit un algorithme numérique simple fondé sur une quadrature de Gauss à 16 points pour calculer la fonction de répartition de l'étendue normale.

Comme le suggèrent Hochberg et al. (1982), une procédure graphique de comparaisons multiples approchant la procédure de comparaisons multiples deux à deux décrite ci-dessus rejette l'hypothèse nulle si, et seulement si

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > q_{\alpha,k} (V_i + V_j) / \sqrt{2}$$

où V_i sont sélectionnés pour minimiser la valeur

$$\sum_{i \neq j} \sum (V_i + V_j - b_{ij})^2$$

où

$$b_{ij} = \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

Comme l'ont illustré Hochberg et al. (1982), la solution à ce problème consiste à utiliser

$$V_i = \frac{(k - 1) \sum_{j \neq i} b_{ij} - \sum_{1 \leq j < l \leq k} b_{jl}}{(k - 1)(k - 2)}$$

Ainsi, un test de l'homogénéité des variances fondé sur cette procédure par approximation rejette l'hypothèse nulle si, et seulement si, au moins deux des intervalles décrits ci-dessous ne se chevauchent pas :

$$\left[S_i \sqrt{c_i \exp(-q_{\alpha,k} V_i / \sqrt{2})}, S_i \sqrt{c_i \exp(q_{\alpha,k} V_i / \sqrt{2})} \right], i = 1, \dots, k$$

La procédure graphique de comparaisons multiples consiste à afficher ces intervalles dans un graphique afin d'identifier visuellement les échantillons dont les intervalles ne se chevauchent pas. En outre, il est possible de déterminer la valeur de p du test global de l'homogénéité de la variance (ou de l'écart type). Dans la section suivante, nous proposons des algorithmes détaillés pour calculer la valeur de p . Toutefois, voici d'abord quelques remarques sur la procédure de comparaisons multiples.

REMARQUE :

1. L'estimateur d'aplatissement regroupé, $\hat{\gamma}_{ij}$, fondé sur la paire d'échantillons (i, j) , aurait pu être remplacé par l'estimateur d'aplatissement regroupé global, portant sur l'ensemble des k échantillons. Bien que cette approche simplifie quelque peu les calculs, des résultats de simulation qui ne sont pas présentés ici indiquent que l'utilisation de $\hat{\gamma}_{ij}$ produit de meilleurs résultats.
2. L'intervalle correspondant à l'échantillon i n'est pas un intervalle de confiance pour l'écart type de la population parent de l'échantillon. Hochberg et al. (1982) appellent cet intervalle "intervalle d'incertitude". Nous l'appelons "intervalle de comparaisons multiples" ou "intervalle CM". Les intervalles de comparaisons multiples permettent seulement de comparer les écarts types ou les variances dans des plans à échantillons multiples.
3. Les intervalles de comparaisons multiples décrits dans cet article permettent uniquement de comparer plus de deux écarts types. S'il n'y a que deux échantillons, il est possible de créer des intervalles de comparaison, mais ces derniers fournissent les mêmes informations que les résultats du test. Il est bien plus intéressant de créer un intervalle de confiance pour le rapport des écarts types, comme celui décrit par Banga et Fox (2013) et disponible avec la commande Variance à 2 échantillons de Minitab.

3. Valeur de p de la méthode graphique de comparaisons multiples

Avant de décrire l'algorithme de calcul de la valeur de p de la méthode graphique de comparaisons multiples, nous calculons dans un premier temps la valeur de p associée à la modification de Bonett (2006) du test de Layard dans des plans à deux échantillons. Nous indiquons ensuite comment appliquer les résultats du plan à deux échantillons à la procédure de comparaisons multiples.

3.1 Valeur de p dans des plans à 2 échantillons

Comme mentionné précédemment, l'ajustement de Bonett (2006) du test de Layard dans les plans à deux échantillons rejette l'hypothèse nulle de l'homogénéité des variances si, et seulement si

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} es$$

ou, autrement dit,

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} se$$

où

$$es = \sqrt{\frac{\hat{\gamma}_{12} - k_1}{n_1 - 1} + \frac{\hat{\gamma}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Bonett a introduit la constante $c_{\alpha/2}$ 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 à petits échantillons. Toutefois, la constante n'a aucun effet dans les plans équilibrés et n'a qu'un effet négligeable dans les plans non équilibrés à grands échantillons.

Ainsi, lorsque le plan est équilibré, la valeur de p du test bilatéral de l'homogénéité des variances se calcule simplement comme suit :

$$P = 2 \Pr(Z > |Z_0|)$$

où

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{es}$$

Si le plan n'est pas équilibré, alors $P = 2 \min(\alpha_L, \alpha_U)$, où α_L est la plus petite solution pour α dans l'équation

$$\exp[\ln(c_{\alpha} S_1^2 / S_2^2) - z_{\alpha} es] = 1 \quad (1)$$

et α_U est la plus petite solution pour α dans l'équation

$$\exp[\ln(c_{\alpha} S_1^2 / S_2^2) + z_{\alpha} es] = 1 \quad (2)$$

Les algorithmes permettant d'obtenir α_L et α_U sont formulés ci-dessous. Le détail mathématique de ces algorithmes est fourni dans la section Annexe.

Soit

$$L(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z es + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Soit également

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{es})}}{2}$$

Les solutions α_L et α_U sont calculées des façons suivantes :

Cas 1 : $n_1 < n_2$

- Calculez z_m comme exprimé dans le résultat ci-dessus et évaluez $L(z_m, n_1, n_2, S_1, S_2)$.
- Si $L(z_m) \leq 0$, calculez le zéro (point d'annulation), z_L , de $L(z, n_1, n_2, S_1, S_2)$ dans l'intervalle, $(-\infty, z_m]$ puis calculez $\alpha_L = \Pr(Z > z_L)$.
- Si $L(z_m) > 0$, la fonction $L(z, n_1, n_2, S_1, S_2)$ n'a pas de zéro. Indiquez $\alpha_L = 0,0$.

Cas 2 : $n_1 > n_2$

- Calculez $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2 / S_2^2$.
- Si $L(0, n_1, n_2, S_1, S_2) \geq 0$, recherchez le zéro, z_o , de $L(z, n_1, n_2, S_1, S_2)$ dans l'intervalle $[0, n_2)$. Sinon, recherchez le zéro z_L dans l'intervalle $(-\infty, 0)$.
- Calculez $\alpha_L = \Pr(Z > z_L)$.

Pour calculer α_U , nous effectuons simplement les étapes ci-dessus en utilisant la fonction $L(z, n_2, n_1, S_2, S_1)$ au lieu de la fonction $L(z, n_1, n_2, S_1, S_2)$.

3.2 Valeur de p des comparaisons multiples graphiques

En supposant que le plan contient k ($k > 2$) échantillons, soit P_{ij} la valeur de p du test associée à une paire d'échantillons (i, j) . Gardez à l'esprit que le test de comparaisons multiples rejette l'hypothèse nulle de l'homogénéité des variances si, et seulement si, au moins deux intervalles de comparaison k ne se chevauchent pas. Ainsi, la valeur de p globale associée au test de comparaisons multiples est égale à

$$P = \min\{P_{ij}, 1 \leq i < j \leq k\}$$

Pour calculer P_{ij} , nous utilisons l'algorithme fourni pour le plan à 2 échantillons, avec

$$es = V_i + V_j$$

où V_i est défini comme précédemment.

Si $n_i \neq n_j$, alors

$$P_{ij} = \min(\alpha_L, \alpha_U)$$

où $\alpha_L = \Pr(Q > z_L \sqrt{2})$, $\alpha_U = \Pr(Q > z_U \sqrt{2})$, z_L est le plus petit zéro de la fonction $L(z, n_i, n_j, S_i, S_j)$, z_U celui de la fonction $L(z, n_j, n_i, S_j, S_i)$ et Q une variable aléatoire comme définie précédemment. L'application de l'algorithme du plan à deux échantillons précédemment décrit à la paire d'échantillons (i, j) permet d'obtenir les quantités z_L et z_U .

Si $n_i = n_j$, alors $P_{ij} = \Pr(Q > |z_o| \sqrt{2})$, où

$$z_o = \frac{\ln S_i^2 - \ln S_j^2}{V_i + V_j}$$

4. Etude par simulation et résultats

Deux grandes études par simulation sont menées pour évaluer les performances du test de comparaisons multiples en tant que test global de l'homogénéité des variances pour de petits échantillons. Toutes les simulations ont été effectuées à l'aide de la version 8 du logiciel Mathematica.

Etude 1

L'objectif de la première étude est d'évaluer et de comparer les propriétés des tests de comparaisons multiples et W_{50} concernant le taux d'erreur de 1ère espèce. Nous comparons les performances des deux tests avec des échantillons générés à l'aide de diverses lois dans trois plans différents : un plan à 3 échantillons, un plan à 4 échantillons et un plan à 6 échantillons. Dans chaque plan, les effectifs d'échantillons varient à raison de 10 à 50 incréments de 10. Les échantillons obéissent aux lois parent suivantes :

- La loi normale
- Des lois de distribution symétriques à queues légères, représentées par la loi de distribution uniforme et la loi bêta, avec les paramètres (3, 3)
- Des lois de distribution symétriques à queues lourdes, représentées par une loi T à 5 degrés de liberté ($t(5)$) et la loi de Laplace
- Des lois de distribution asymétriques à queues lourdes, représentées par la loi exponentielle et des lois du Khi deux à 1 et 5 degrés de liberté ($\chi^2(1)$) et ($\chi^2(5)$), respectivement
- Une loi normale contaminée (CN(0,9, 3)) pour laquelle 90 % des observations sont issues de la loi normale standard et les 10 % restants d'une loi normale ayant pour moyenne 0 et pour écart type 3.

Chaque simulation porte sur 10 000 répliques d'échantillons. Le niveau α nominal cible est de 0,05. L'erreur de simulation est d'environ 0,002. Les seuils de signification simulés pour chaque test sont indiqués dans le tableau 1.

Tableau 1 Comparaison des seuils de signification simulés ($\alpha = 0,05$)

Description	Loi de distribution [aplatissement]	n_i	$k = 3$		$k = 4$		$k = 6$	
			CM	W_{50}	CM	W_{50}	CM	W_{50}
Normale	Normale [3,0]	10	,038	,033	,038	,031	,036	,029
		20	,039	,038	,040	,038	,041	,033
		30	,043	,041	,044	,038	,046	,039
		40	,046	,043	,046	,041	,048	,041
		50	,046	,046	,046	,044	,052	,047

Description	Loi de distribution [aplatissement]	n_i	$k = 3$		$k = 4$		$k = 6$	
			CM	W_{50}	CM	W_{50}	CM	W_{50}
Symétriques à queues légères	Uniforme [1,8]	10	,029	,029	,025	,024	,023	,020
		20	,028	,026	,030	,026	,028	,023
		30	,037	,035	,034	,032	,034	,030
		40	,038	,037	,037	,037	,035	,033
		50	,041	,041	,036	,036	,036	,036
	Bêta(3, 3) [2,5]	10	,031	,032	,031	,029	,031	,025
		20	,035	,031	,036	,027	,037	,026
		30	,041	,035	,037	,034	,037	,032
		40	,040	,036	,039	,035	,040	,033
		50	,044	,039	,044	,037	,044	,035
Symétriques à queues lourdes	Laplace [6,0]	10	,056	,038	,063	,041	,071	,039
		20	,054	,044	,058	,043	,059	,041
		30	,051	,042	,053	,043	,052	,044
		40	,048	,045	,048	,045	,048	,046
		50	,045	,045	,051	,046	,049	,047
	$t(5)$ [9,0]	10	,042	,032	,044	,031	,042	,031
		20	,043	,039	,045	,038	,045	,040
		30	,039	,040	,040	,040	,041	,040
		40	,041	,042	,040	,041	,039	,038
		50	,040	,050	,039	,046	,038	,046
Asymétriques à queues lourdes	$\chi^2(5)$ [5,4]	10	,040	,039	,046	,040	,048	,039
		20	,040	,043	,040	,040	,042	,039
		30	,039	,047	,042	,044	,043	,042
		40	,040	,046	,041	,044	,039	,042
		50	,037	,047	,038	,047	,040	,048
	Exponentielle [9,0]	10	,063	,051	,073	,049	,076	,048
		20	,051	,049	,053	,048	,057	,046
		30	,042	,048	,046	,051	,049	,049
		40	,034	,050	,038	,046	,037	,049
		50	,033	,045	,037	,047	,038	,046

Description	Loi de distribution [aplatissement]	n_i	$k = 3$		$k = 4$		$k = 6$	
			CM	W_{50}	CM	W_{50}	CM	W_{50}
	$\chi^2(1)$ [15,0]	10	,084	,048	,098	,050	,118	,050
		20	,053	,046	,060	,047	,068	,046
		30	,041	,041	,045	,045	,050	,047
		40	,044	,049	,046	,047	,045	,047
		50	,038	,050	,037	,049	,040	,049
Normale contaminée	CN(0,9, 3) [8,3]	10	,020	,016	,018	,012	,016	,010
		20	,014	,015	,012	,013	,008	,007
		30	,012	,014	,010	,011	,007	,008
		40	,009	,017	,009	,014	,006	,008
		50	,009	,016	,007	,012	,006	,009

Les résultats indiquent que les deux tests donnent de bons résultats pour la plupart des lois. La plupart des seuils de signification simulés sont proches de la valeur cible de 0,05. Toutefois, les seuils de signification simulés pour les deux tests ont tendance à être prudents (inférieurs à 0,05) lorsque de petits échantillons sont générés avec des lois normales et symétriques à queues légères. Pour ces lois, les seuils de signification simulés du test de comparaisons multiples sont plus proches du seuil de signification cible que ceux du test W_{50} .

Lorsque de petits échantillons sont générés avec des lois à queues lourdes, le test W_{50} a tendance à être prudent et le test de comparaisons multiples, libéral. Le test de comparaisons multiples est encore plus libéral lorsque de petits échantillons sont générés avec des lois très asymétriques. Par exemple, lorsque des échantillons ayant des effectifs de 10 sont créés à l'aide d'une loi du Khi deux à 1 degré de liberté, les seuils de signification simulés pour le test de comparaisons multiples sont de 0,084, 0,098 et 0,118 pour les plans à 3, 4 et 6 échantillons, respectivement.

Les valeurs aberrantes influent sur les deux tests. Les seuils de signification de la distribution normale contaminée sont extrêmement prudents, même lorsque les échantillons sont aussi grands que 50.

Etude 2

Dans la deuxième étude, nous évaluons et comparons les propriétés des deux procédures concernant le taux d'erreur de 2ème espèce (puissance) dans un plan à 4 échantillons. Pour cette étude, nous utilisons les mêmes échantillons que ceux utilisés dans l'étude 1 pour un effectif de 20 et pour $k = 4$. Les observations sont mises à l'échelle par un facteur de 1, 2, 3 ou 4. Par exemple, dans la condition écrite 1:1:4:4, les observations des échantillons 1 et 2 sont les mêmes que celles utilisées dans l'étude 1. Les observations des échantillons 3 et 4 sont mises à l'échelle par un facteur de 4.

Nous incluons la condition 1:1:1:1 comme base de comparaison. Notez que les résultats obtenus avec cette condition sont les mêmes que ceux indiqués dans l'étude 1 pour des échantillons ayant des effectifs de 20 et pour $k = 4$. Nous optons pour des échantillons ayant des effectifs de 20, car les résultats de l'étude 1 suggèrent que, pour les deux tests, les échantillons ayant des effectifs de 20 produisent des seuils de signification proches du seuil cible pour la plupart des lois.

Les niveaux de puissance simulée de ces expériences sont calculés comme la proportion de répliques d'échantillons conduisant à un rejet de l'hypothèse nulle d'homogénéité des variances.

Les résultats sont présentés dans le tableau 2.

Tableau 2 Comparaison des niveaux de puissance simulée ($\alpha = 0,05$)

Description	Loi de distribution	Rapport des écarts types							
		1:1:1:1		1:1:2:2		1:2:3:4		1:1:4:4	
		CM	W_{50}	CM	W_{50}	CM	W_{50}	CM	W_{50}
	Normale	,040	,038	,846	,853	,998	,994	1,000	1,000
Symétriques à queues légères	Uniforme	,030	,026	,985	,962	1,000	,999	1,000	1,000
	Bêta(3, 3)	,036	,027	,938	,916	1,000	,999	1,000	1,000
Symétriques à queues lourdes	Laplace	,058	,043	,597	,629	,931	,921	,996	,998
	$t(5)$,045	,038	,657	,703	,952	,949	,997	,998
Asymétriques à queues lourdes	$\chi^2(5)$,040	,040	,625	,704	,949	,949	,996	,999
	Exponentielle	,053	,048	,431	,507	,804	,779	,963	,978
	$\chi^2(1)$,060	,047	,298	,291	,602	,504	,838	,824
Contaminée	CN(0,9, 3)	,012	,013	,499	,612	,889	,917	,989	,998

Les résultats suggèrent que les propriétés du test W_{50} et du test de comparaisons multiples concernant le taux d'erreur de 2ème espèce (puissance) sont similaires. En général, les niveaux de puissance simulée obtenus avec les deux tests sont du même ordre de grandeur. La puissance des deux tests ne diffère de plus de 0,1 que dans un seul cas.

Les niveaux de puissance simulée pour le test de comparaisons multiples sont légèrement meilleurs que ceux du test W_{50} lorsque les échantillons sont créés à l'aide de lois symétriques à queues légères ou modérées. En revanche, le test W_{50} semble être légèrement plus puissant que le test de comparaisons multiples lorsque les échantillons sont créés à l'aide de lois à queues lourdes.

5. Exemple

Dans cette section, nous appliquons la procédure graphique de comparaisons multiples et le test W_{50} à un ensemble de données fourni par Ott et al. (2010), page 397. Les données sont décrites de la façon suivante :

Une entreprise métallurgique spécialisée dans le coulage dispose de plusieurs fours dans lesquels les matières premières sont chauffées avant d'être versées dans un moule en cire. Il est essentiel que ces métaux soient chauffés à une température précise qui présente peu de variation. Trois fours sont choisis de façon aléatoire et leurs températures exactes sont observées (°C) lors de 10 opérations de chauffage successives. Les données collectées sont les suivantes :

Four 1	1670,87	1670,88	1671,51	1672,01	1669,63	1670,95	1668,70	1671,86	1669,12	1672,52
Four 2	1669,16	1669,60	1669,76	1669,18	1671,92	1669,69	1669,45	1669,35	1671,89	1673,45
Four 3	1673,08	1672,75	1675,14	1674,94	1671,33	1660,38	1679,94	1660,51	1668,78	1664,32

La figure 1 présente les boîtes à moustaches des températures de chaque four. Les boîtes à moustaches indiquent qu'il n'existe pas de valeurs aberrantes dans les températures enregistrées et que la variabilité de la température du four 3 diffère de celle des fours 1 et 2.

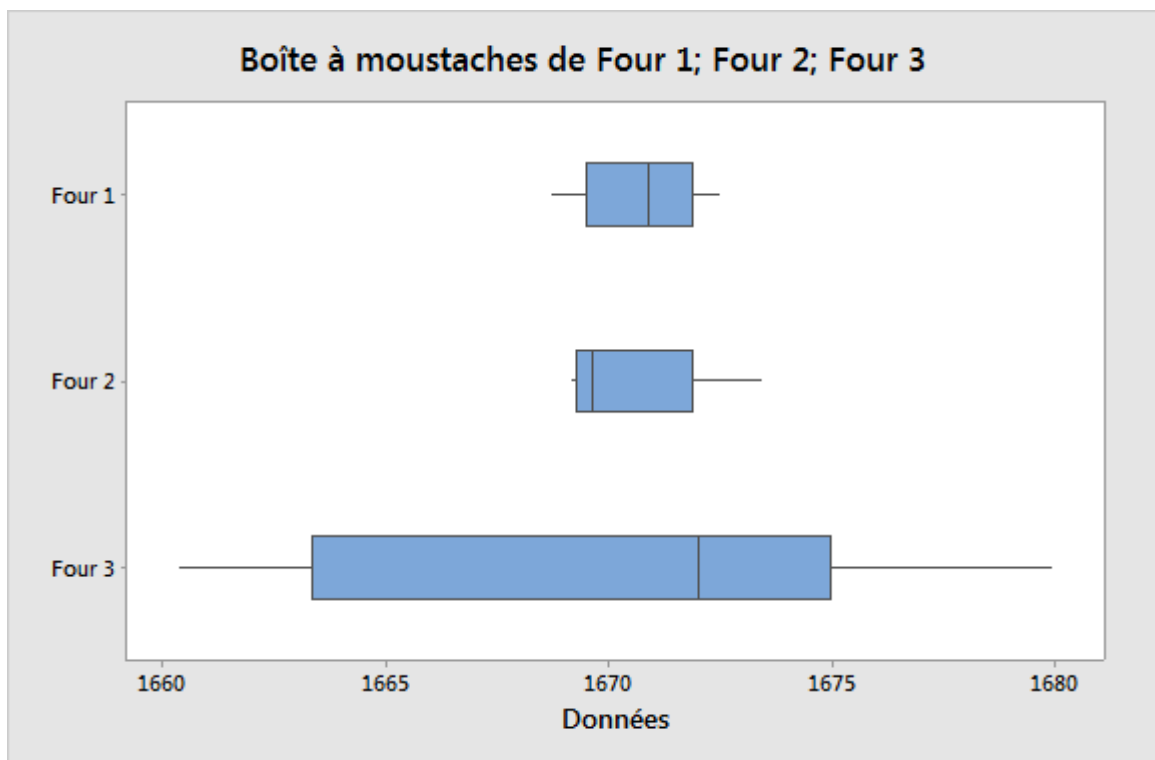


Figure 1 Boîtes à moustaches de la température du four (°C)

La figure 2 indique les intervalles de comparaisons multiples pour les mêmes données, ainsi que les résultats du test global de comparaisons multiples et du test W_{50} , appelé test de Levene dans la légende. Les valeurs de p significatives pour les deux tests indiquent que la variabilité de la température diffère entre les trois fours. Le fait que les intervalles de comparaisons multiples ne se chevauchent pas confirme que la variabilité du four 3 diffère de celle des autres fours. Les intervalles de comparaisons multiples sont (0,896, 2,378), (1,072, 2,760) et (4,366, 12,787) pour les fours 1, 2 et 3, respectivement.

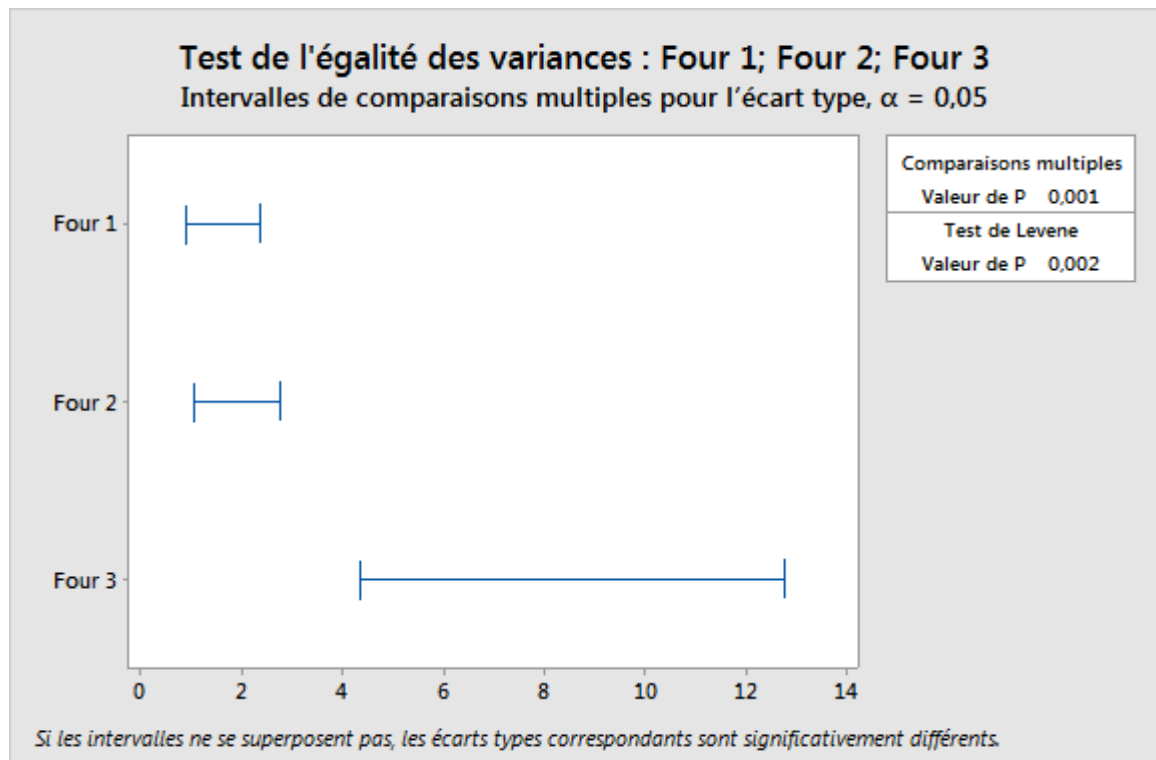


Figure 2 Intervalles de comparaisons multiples et valeurs de p des tests de comparaisons multiples et W_{50} (test de Levene)

6. Conclusion

Dans l'ensemble, les résultats de la simulation indiquent que pour des plans comportant plusieurs petits échantillons, les performances du test de comparaisons multiples sont semblables à celles du test W_{50} . Le test de comparaisons multiples est légèrement plus adapté pour des lois symétriques ou presque symétriques à queues légères ou modérées. En revanche, le test W_{50} est sans doute préférable lorsque les données obéissent à des lois très asymétriques ou à queues lourdes. Un avantage certain de la procédure de comparaisons multiples est qu'elle fournit un outil visuel pratique pour identifier les échantillons présentant des écarts types ou des variances différentes lorsque le test global de l'homogénéité des écarts types est significatif. La procédure graphique de comparaisons multiples est disponible dans la version 17 de Minitab.

7. Annexe

L'ajustement de Bonett (2006) du test de Layard dans les plans à deux échantillons rejette l'hypothèse nulle de l'homogénéité des variances si, et seulement si,

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} es$$

ou, autrement dit,

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} es$$

où

$$es = \sqrt{\frac{\hat{y}_{12} - k_1}{n_1 - 1} + \frac{\hat{y}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Ainsi, si le plan est équilibré, alors $c_{\alpha/2} = 1$ et la valeur de p du test est simplement obtenue ainsi :

$$P = 2 \Pr(Z > |Z_0|)$$

où

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{es}$$

Si le plan est non équilibré, alors $P = 2 \min(\alpha_L, \alpha_U)$ où

α_L est la plus petite solution pour α dans l'équation

$$\exp[\ln(c_\alpha S_1^2 / S_2^2) - z_\alpha es] = 1 \quad (1)$$

et α_U est la plus petite solution pour α dans l'équation

$$\exp[\ln(c_\alpha S_1^2 / S_2^2) + z_\alpha es] = 1 \quad (2)$$

La méthode permettant de résoudre ces équations pour α consiste dans un premier temps à résoudre les équations pour $z \equiv z_\alpha$, afin d'obtenir ensuite $\alpha = \Pr(Z > z)$ où la variable aléatoire Z est distribuée selon la loi normale standard. Avant d'en venir à la résolution de ces équations, notons que l'équation (1) peut être exprimée différemment sous la forme $L(z) = 0$, où

$$L(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z es + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

De même, l'équation (2) équivaut à l'équation $U(z) = 0$, où

$$U(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} + z es + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Notons que $L(z, n_2, n_1, S_2, S_1) = -U(z, n_1, n_2, S_1, S_2)$. Par conséquent, il ne faut rechercher les zéros que d'une seule des deux fonctions.

L'algorithme permettant de résoudre l'équation (1) ou (2) est déduit du résultat suivant :

Résultat

Soit n_1, n_2, S_1 et S_2 des valeurs données fixes. Dans des plans non équilibrés, la fonction $L(z, n_1, n_2, S_1, S_2)$ admet au moins deux zéros.

4. Si $n_1 < n_2$, alors $L(z, n_1, n_2, S_1, S_2)$ est convexe : il satisfait $L(-\infty, n_1, n_2, S_1, S_2) = L(n_1, n_1, n_2, S_1, S_2) = +\infty$ et atteint sa valeur minimale en

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{es})}}{2}$$

Ainsi, si $L(z_m, n_1, n_2, S_1, S_2) \leq 0$, il existe deux zéros : un dans l'intervalle $(-\infty, z_m]$ et l'autre dans l'intervalle $[z_m, n_1)$. En revanche, si $L(z_m, n_1, n_2, S_1, S_2) > 0$, la fonction $L(z, n_1, n_2, S_1, S_2)$ n'admet aucun de zéro.

5. Si $n_1 > n_2$, alors $L(z, n_1, n_2, S_1, S_2)$ décroît de façon monotone entre $+\infty$ et $-\infty$ et n'admet donc qu'un seul zéro. Si $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2 \geq 0$, le zéro se trouve dans l'intervalle $[0, n_2)$; sinon, il est compris dans l'intervalle $(-\infty, 0)$.

Démonstration

Soit $L(z) \equiv L(z, n_1, n_2, S_1, S_2)$

Nous souhaitons dans un premier temps démontrer que, si $n_1 < n_2$ alors $L(z)$ est convexe et atteint sa valeur minimale en

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{es})}}{2}$$

Comme défini précédemment

$$L(z) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z es + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Nous avons donc $\lim_{z \rightarrow -\infty} L(z) = +\infty$ et

$$\lim_{z \rightarrow \min(n_1, n_2)} L(z) = \begin{cases} +\infty & \text{si } n_1 < n_2 \\ -\infty & \text{si } n_2 < n_1 \end{cases}$$

Notez également que la dérivée de $L(z)$ satisfait

$$-\frac{(n_1 - z)(n_2 - z)}{es} L'(z) = z^2 - (n_1 + n_2)z + n_1 n_2 + \frac{n_1 - n_2}{es}$$

Soit

$$Q(z) = -\frac{(n_1 - z)(n_2 - z)}{es} L'(z)$$

Si $n_1 < n_2$, la fonction quadratique $Q(z)$ a deux zéros, exprimés de la façon suivante

$$z_1 = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{es})}}{2}$$

et

$$z_2 = \frac{n_1 + n_2 + \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{es})}}{2}$$

Etant donné que $Q(n_1) = \frac{n_1 - n_2}{es} < 0$, $z_1 < n_1 = \min(n_1, n_2) < z_2$ de telle sorte que $Q(z) > 0$ pour z dans $(-\infty, z_1)$ et que $Q(z) < 0$ pour z dans (z_1, n_1) . Ainsi, $L'(z) < 0$ lorsque z appartient à $(-\infty, z_1)$ et $L'(z) > 0$ lorsque z appartient à (z_1, n_1) . Par conséquent, $L(z)$ est convexe dans le domaine $(-\infty, \min(n_1, n_2))$ et atteint sa valeur minimale lorsque $z_1 \equiv z_m$.

Si $n_1 > n_2$, deux cas de figure sont possibles : soit $n_1 - n_2 > 4/es$, soit

$0 < n_1 - n_2 < 4/es$. Dans le premier cas, z_1 et z_2 sont les zéros de $Q(z)$, tel que $n_2 = \min(n_1, n_2) < z_1 < z_2$. (Cela est dû au fait que $n_2 - \frac{z_1 + z_2}{2} = \frac{n_2 - n_1}{2} < 0$). Ainsi, $Q(z) > 0$ lorsque z appartient au domaine $(-\infty, \min(n_1, n_2))$. Dans le second cas, $Q(z)$ n'admet aucun zéro, et donc $Q(z) > 0$ sur le domaine.

Ainsi, si $n_1 > n_2$, alors $L'(z) < 0$, et donc $L(z)$ décroît de façon monotone entre $+\infty$ et $-\infty$.

8. Références

Banga, S. J. et Fox, G. D. (2013), Sur l'intervalle de confiance robuste de Bonett pour le rapport des écarts types, Sous presse.

Barnard, J. (1978), Probability Integral of the Normal Range, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, 197–198.

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

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.

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

Kramer, C. Y. (1956), Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications, *Biometrics*, 12, 307–310.

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.

Miller, R. G. (1968), Jackknifing Variances, *Annals of Mathematical Statistics*, 39, 567–582.

Nakayama, M. K. (2009), Asymptotically Valid Single-Stage Multiple-Comparison Procedures, *Journal of Statistical Planning and Inference*, 139, 1348–1356.

Ott, R. L. et Longnecker, M. (2010), *An introduction to Statistical Methods and Data Analysis*, sixième édition, Brooks/Cole, Cengage Learning.

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

Stoline, M. R. (1981), The Status of Multiple of Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs, *The American Statistician*, 35, 134–141.

Tukey, J. W. (1953), *The Problem of Multiple Comparisons*. Monographie mimeographiée.

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

© 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.