Au delà du test de signification ou l'inférence statistique sans tables (à la suite d'Alain Morineau)

Bruno Lecoutre

ERIS, Laboratoire de Mathématiques Raphaël Salem
UMR 6085 C.N.R.S. et Université de Rouen
Mathématiques Site Colbert, 76821 Mont-Saint-Aignan Cedex
bruno.lecoutre@univ-rouen.fr
Internet: http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris.htm

Alain Morineau (1995) nous rappelle que des réflexes très simples nous permettent de "faire de l'inférence statistique sans tables". A l'ère informatique, il est particulièrement important de pouvoir ainsi disposer de repères qui nous permettent un regard critique sur des avalanches de résultats. Mais plus encore, la simplicité de son résultat nous amène à réfléchir sur la pertinence de la procédure: répond-elle véritablement aux questions que nous nous posons? Ainsi pour une corrélation le résultat est en gros significatif si la corrélation observée est supérieure à $2/\sqrt{n}$. Pour $n = 900$, une corrélation observée de 0.07 sera donc significative, tandis que pour $n = 50$, une corrélation de 0.27 ne sera pas significative... En fait nous n'avons pas seulement (ou même nous n'avons pas...) besoin d'une procédure de décision brutale, qui ne concerne que la valeur zéro et ne nous renseigne pas sur l'importance réelle de la corrélation. Mais nous devons aussi pouvoir "tester" d'autres valeurs, et plus simplement obtenir une "fourchette" qui nous permette d'apprécier réellement l'information apportée par les données.

Nous allons rappeler que, dans les cas les plus courants de traitements de données numériques, il est immédiat de passer du test usuel à cette fourchette. Bien entendu il faudra justifier et interpréter celle-ci; on pourra se réjouir de savoir qu'elle peut être regardée comme un intervalle de confiance (*fréquentiste*), comme un intervalle *fiduciaire*, ou comme un intervalle de crédibilité *bayésien* standard. Dans la suite nous l'appellerons simplement "intervalle", laissant le lecteur libre de choisir son *cadre de justification et d'interprétation*. [Pour en savoir plus: Rouanet *et al.*, 1991, 2000].

Au delà du test de Student

Considérons par exemple la situation usuelle de la comparaison de deux moyennes de groupes indépendants. Si nous connaissons la différence observée d (que nous supposerons non nulle) et la valeur t prise par la statistique du test de Student usuel de l'hypothèse nulle d'une différence *vraie* $\delta = 0$, il est immédiat de procéder au test de toute autre hypothèse $\delta = \delta_0$. Ce test est significatif (au seuil bilatéral 0.05) si la différence $\delta - \delta_0$ est en valeur absolue supérieure à $2d/t$. Il est non moins immédiat d'en déduire les bornes (approchées) de "l'intervalle à 95%" pour la différence vraie δ:

$$d - 2d/t \quad \text{et} \quad d + 2d/t$$

Le lien entre les deux procédures est assuré par le fait que l'intervalle est l'ensemble des valeurs d_0 pour lesquelles le test est non-signicatif. Ces résultats sont en fait à valeur générale pour toute inférence sur une combinaison linéaire de moyennes de groupes indépendants et/ou appariés sous les modèles usuels. Si on dispose des résultats d'une analyse de variance (ANOVA) on remplacera t par la racine carrée du rapport F (et d par sa valeur absolue s'il y a lieu). Pour plus de précision (mais un peu moins de simplicité) on peut remplacer 2 par $1.96\sqrt{q/(q-2)}$. Nous pouvons ainsi immédiatement prolonger les tests de signification usuels et en particulier résoudre la plupart des ambiguïtés que leur usage soulève dans les publications expérimentales. Par exemple nous trouvons dans un article (en anglais) d'une revue internationale le tableau de moyennes suivant:

	treatment 1	treatment 2	
age 1	4.71	5.25	5.08
age 2	4.83	5.77	5.30
	4.77	5.51	

Les commentaires de ces résultats font appel à des expressions consacrées, "*The only significant effect is a main effect of treatment* (F[1,36] = 6.39, p = 0.016)*, reflecting a substantial improvement in accuracy*" et encore, "*Clearly, there is no evidence* (F[1,36] = 0.47, p = 0.50) *of an interaction*".

Le lecteur peu au fait de la rhétorique qui accompagne l'usage des tests pourrait en déduire que l'on a démontré que la différence entre les traitements est importante, tandis que l'effet d'interaction est, sinon nul, du moins relativement négligeable. Il est facile de voir qu'il n'en est rien. En effet, pour la comparaison des deux traitements la différence observée est $d = +0.74$, d'où $2|d|/\sqrt{F} = 0.59$ et l'intervalle $[+0.15, +1.33]$, tandis que pour l'interaction la différence des différences observée est 0.40, d'où $2|d|/\sqrt{F} = 1.17$ et l'intervalle $[-0.77, +1.57]$. (La précision avec laquelle sont effectués les calculs est largement suffisante pour les besoins de la démonstration). [Pour en savoir plus: Lecoutre, 1996; Lecoutre et Poitevineau, 1996]

Au delà du coefficient de corrélation

Dans ce cas la situation est plus complexe. En effet le résultat simple dans le cas de l'hypothèse nulle d'une corrélation vraie ρ égale à zéro ne s'applique pas à une autre valeur. Si $\rho = \rho_0$ la distribution de r est généralement très asymétrique et sa variance dépend de ρ_0. Mais on connaît, au moins depuis Fisher en 1915, le moyen de surmonter cette difficulté. Il s'agit de considérer les quantités

$$z = tanh^{-1}\, r \quad (\text{ou } Argth\, r) \quad \text{et} \quad \zeta_0 = tanh^{-1}\, r_0$$

qui à notre époque peuvent être obtenues immédiatement avec une calculatrice. La distribution de z est bien approximée (d'autant plus que n est grand et $|\rho_0|$ pas trop proche de 1) par une distribution normale de moyenne ζ_0 et de variance $1/(n-3)$, d'où des procédures aussi simples à la transformation près que dans le cas $r_0 = 0$.

Pour tester l'hypothèse nulle $\rho = \rho_0$ on compare la différence $|z - z_0|$ à $2/\sqrt{n-3}$, et les bornes de "l'intervalle à 95%" pour ρ sont

$$tanh(z - 2/\sqrt{n-3}) \quad \text{et} \quad tanh(z + 2/\sqrt{n-3})$$

(pour plus de précision on remplacera 2 par 1.96)

Dans l'exemple de Morineau ($r = 0.18$) nous obtenons $z = 0.1820$ (pour r proche de 0 on a $z \approx r$) d'où par exemple les intervalles $[+0.11, +0.24]$ pour $n = 900$ et $[-0.11, +0.44]$ pour $n = 50$. [Pour en savoir plus: Fisher, 1990]

Références

Fisher R.A. (1990) – *Statistical Methods, Experimental Design and Scientific Inference*. Oxford University Press (réédition).

Lecoutre B. (1996) – *Traitement statistique des données expérimentales: Des pratiques traditionnelles aux pratiques bayésiennes - Avec programmes Windows® par B. Lecoutre et*

J. Poitevineau. Saint-Mandé: C.I.S.I.A. [Ces programmes sont également disponibles sur Internet à l'adresse: http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/pac.htm].

Morineau A. (1995) - Tester une corrélation sans table... *La Revue de Modulad* 16, 41.

Rouanet H., Lecoutre M.-P., Bert M.-C., Lecoutre B., Bernard J.-M.(1991) – *L'inférence statistique dans la démarche du chercheur*. Publications Universitaires Européennes. Berne: Peter Lang.

Rouanet J.-M., Bernard, J.M., Bert, M.-C., Lecoutre, B., Lecoutre, M.-P., Le Roux, B. – *New ways in statistical methodology: From significance tests to Bayesian inference* (2ème édition), 29-64, Peter Lang, Berne.