

O28-Traitement statistique des données appariées. Application à la comparaison de données réelles et imaginées.

Stéphane Champely, Christian Collet & Aymeric Guillot

Centre de Recherche et d'Innovation sur le Sport, EA647, Université de Lyon, Université Claude Bernard Lyon 1, France.
champely@univ-lyon1.fr

Cette communication vise à synthétiser des propositions peu connues, bien que parfois anciennes, sur le traitement statistique des données appariées, un des dispositifs expérimentaux les plus employés en sciences du sport. Le cas particulier envisagé est celui de la comparaison de performances réelles et imaginées dans quatre groupes (deux activités : ski et équitation et deux niveaux : débutant et expert). Au-delà du test apparié de Student de comparaison de moyennes, il est possible d'utiliser de représentations graphiques spécifiques, de comparer les variances et enfin d'utiliser des versions « modernes » de ces procédures inférentielles car la robustesse des méthodes classiques n'est plus si évidente à la lueur d'études récentes.

Mots-clefs : Statistique, Données appariées, Graphiques, Robustesse, Données réelles et imaginées.

INTRODUCTION

Afin de comparer efficacement deux conditions expérimentales, le dispositif apparié consiste à appliquer les deux « traitements » à des paires d'unités statistiques préalablement constituées pour leur « ressemblance ». Cette paire peut être un patient et un contrôle lui correspondant, un même individu mesuré longitudinalement (avant-après) ou sous deux conditions différentes (mesures répétées classiques). L'appariement permet parfois d'augmenter considérablement la puissance statistique de l'étude et donc de réduire le nombre de sujets. Il est très couramment employé dans les sciences du sport.

En pratique, l'analyse statistique de telles données reste le plus souvent limitée à un test de Student apparié ou de façon équivalente une analyse de variance à mesures répétées dont l'objectif est de détecter une différence de localisation entre les échantillons. Dans le contexte particulier des études de fiabilité *test-retest* (Atkinson & Nevill, 1998), d'autres propositions avec des objectifs bien spécifiques et un peu différents visant à déterminer l'accord (*agreement*) entre les deux mesures ont été avancées : divers coefficients de corrélation, plusieurs statistiques de dispersion intra-individu...

Si de nouvelles méthodes pour traiter les données appariées ont émergé de la littérature statistique, c'est de façon éparse dans le temps et les revues et fragmentée sur des aspects spécifiques. Et bien que la situation appariée soit un problème de base de la pratique statistique, ces innovations ne sont finalement pas reprises dans des ouvrages d'introduction et quasiment absentes des ouvrages plus techniques.

L'objectif de cette communication est de synthétiser ces méthodes en les développant parfois, afin de proposer une gamme complète d'outils statistiques pour étudier des données appariées. Il sera successivement question d'analyse graphique, de tests de variance et de tests robustes.

LES DONNÉES : MESURES DE PERFORMANCES RÉELLES ET IMAGINÉES

Des mesures réelles de performance en compétition – durée de parcours en secondes – ont été prises pour des skieurs experts ($n=12$) et débutants ($n=8$) ainsi que des cavaliers experts ($n=8$) et débutants ($n=8$). De façon parallèle, des mesures imaginées ont été collectées après une reconnaissance du parcours. Nous noterons les données réelles X et les données imaginées Y . Des tests t appariés montrent des différences significatives chez les débutants en ski ($t=3.0$, $df=7$, $p=0.019$) et en équitation ($t=4.1$, $df=7$, $p=0.004$), chez les cavaliers experts ($t=6.2$, $df=7$, $p<0.001$), mais pas en ce qui concerne les skieurs experts ($t=-0.6$, $df=11$, $p=0.536$).

LES REPRÉSENTATIONS GRAPHIQUES

La visualisation est essentielle en statistique, pour communiquer bien entendu les résultats, mais aussi afin d'explorer la structure des données. Quel graphique employer pour des données appariées ? Les logiciels usuels ne proposent généralement rien de véritablement

adapté et on doit se contenter soit de deux boîtes à moustaches simultanées (où l'appariement disparaît...) soit d'un nuage de points avec des défauts que nous soulignerons soit, après une petite manipulation, d'une représentation univariée des différences $D=Y-X$.

Heureusement, quatre types de graphiques plus adéquats sont envisageables : le graphique en coordonnées parallèles, le graphique en lignes parallèles, le graphique différence versus moyenne et le *sliced squared plot*. Ces représentations permettent en particulier de détecter des écarts aux hypothèses des modèles probabilistes envisagés.

LES TESTS DE COMPARAISON DE VARIANCES

Si le test de Student compare la moyenne des deux échantillons, il est aussi intéressant de se pencher sur leurs variances respectives. Le test de Pitman-Morgan (1939) permet de le faire pour des données corrélées et repose sur une astuce simple : tester la corrélation entre la différence D et la somme $S=X+Y$. Une corrélation significative revient à une différence entre les deux variances. Ce test a été prolongé par Bradley et Blackwood (1989) qui proposent un test simultané des moyennes et des variances basé sur la régression de D sur S . Nous ferons une proposition inédite afin de « lire ces tests » sur certaines des représentations graphiques.

LES VERSIONS ROBUSTES DES TESTS DE COMPARAISON

Une méthode statistique est robuste si ses performances sont peu altérées par des écarts aux hypothèses de référence (a-normalité, hétéroscédasticité, points extrêmes...) S'il est depuis longtemps reconnu que les tests comparant les variances présentent de sérieux problèmes de robustesse, une certaine confiance a longtemps régné dans les qualités de ceux comparant les moyennes. Il semble toutefois que leurs performances, en particulier en termes de puissance et de couverture d'intervalle de confiance, ne soient en fait pas si brillantes pour des écarts qui se rencontrent couramment avec des données réelles (Wilcox, 2005).

Si la situation à deux échantillons indépendants a fait l'objet de multiples études et propositions, c'est moins le cas pour deux échantillons appariés. On peut cependant trouver des méthodes robustes permettant d'étudier les échantillons corrélés à la fois en localisation et en dispersion (Welch & Gutierrez, 1988 ; Grambsch, 1994). Nous montrerons sur les données de performances considérées les résultats obtenus avec de telles méthodes ainsi que des simulations permettant de comparer les qualités des techniques classiques et robustes. Nous discuterons également du test de Wilcoxon apparié (Munzel, 1999) qui est souvent présenté comme la solution au problème de distributions non normales.

BIBLIOGRAPHIE

- Atkinson, G. & Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medecine*, 26, 217-238.
- Bradley, E.L. & Blackwood, L.G. (1989). Comparing paired data: A simultaneous test of means and variances. *The American Statistician*, 43, 234-235.
- Grambsch, P.M. (1994). Simple robust tests for scale differences in paired data. *Biometrika*, 81, 359-372.
- Munzel, U. (1999). Nonparametric methods for paired samples. *Statistica Neerlandica*, 53, 277-286.
- Pitman, E.J.G. (1939). A note on normal correlation. *Biometrika*, 31, 9-12.
- Welch, W.J. & Gutierrez, L.G. (1988). Robust permutation tests for matched paired designs. *Journal of the American Statistical Association*, 83, 450-455.
- Wilcox, R.R. (2005). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.