

# L'approche par comparaison de modèles avec R2STATS dans l'enseignement des statistiques en sciences humaines

Yvonnick Noël<sup>a</sup>

<sup>a</sup>Département de Psychologie  
Université Européenne de Bretagne, Rennes 2  
Place du Recteur Henri Le Moal, 35043 Rennes Cedex  
yvonnick.noel@uhb.fr

**Mots clefs** : Interfaces graphiques pour R, GLM(M), facteur de Bayes, comparaison de modèles.

L'enseignement des statistiques aux étudiants en sciences humaines pose des problèmes spécifiques. Le public concerné est hétérogène et peu de nos étudiants montrent un goût affirmé pour la ligne de commande. Nous expérimentons depuis plusieurs années un enseignement basé sur la comparaison de modèles à l'aide d'interfaces graphiques pour R. La première, nommée R2STATS (Noël, 2012a) est une interface aux fonctions `glm()` de la librairie `base` et `glmer()` de la librairie `lme4`. Elle permet de construire facilement une séquence de GLM (ou de GLMM) emboîtés, d'en obtenir une représentation graphique automatique et de les comparer entre eux par des statistiques traditionnelles ( $F$  de Fisher,  $\chi^2$  sur la réduction de la déviance). La deuxième, nommée `AtelierR` (Noël, 2012b) inclut des outils de décision bayésiens plus récents (le facteur de Bayes) et des fonctions avancées de recherche automatique du meilleur modèle.

Ces interfaces encouragent l'approche qui consiste à traduire une question psychologique en une mise en concurrence de plusieurs modèles possibles. Dans cette approche, le point de départ est toujours le choix d'un modèle de distribution (conditionnel) sur la variable dépendante. C'est ce choix, argumenté théoriquement ou empiriquement (par un test d'ajustement), qui définit le type de problème statistique. Par exemple, la comparaison des résultats obtenus dans deux conditions expérimentales différentes ne se traduit par une comparaison de moyennes que dans le cadre d'une hypothèse de normalité et d'homogénéité des variances, car ce sont ces deux conditions qui réduisent la comparaison des modèles à une simple comparaison de paramètres de moyennes. Sous R2STATS, l'utilisateur définit d'abord son modèle de distribution (Figure 1), et les statistiques usuelles apparaissent dans le processus d'estimation des paramètres du modèle choisi ( $T$  de Student sur un coefficient), ou dans une comparaison de modèles concurrents ( $F$  de Fisher,  $\chi^2$  sur la réduction de la déviance).

L'interface `AtelierR` (Figure 2) implémente la même chose, dans une perspective bayésienne, en cherchant automatiquement le modèle le plus probablement vrai, sur un ensemble supposé exhaustif, pour les modèles binomiaux, multinomiaux et gaussiens (Noël, 2012b, 2013).

## Références

- [1] Noël, Y. (2013). *Psychologie statistique avec R*, coll. Pratique R, Paris : Springer.
- [2] Noël, Y. (2012a). R2STATS: A GTK GUI for fitting and comparing GLM and GLMM in R. R package version 0.68-31. <http://CRAN.R-project.org/package=R2STATS>
- [3] Noël, Y. (2012b). `AtelierR`: A GTK GUI for teaching basic concepts in Bayesian statistical inference. R package version 0.23. <http://CRAN.R-project.org/package=AtelierR>

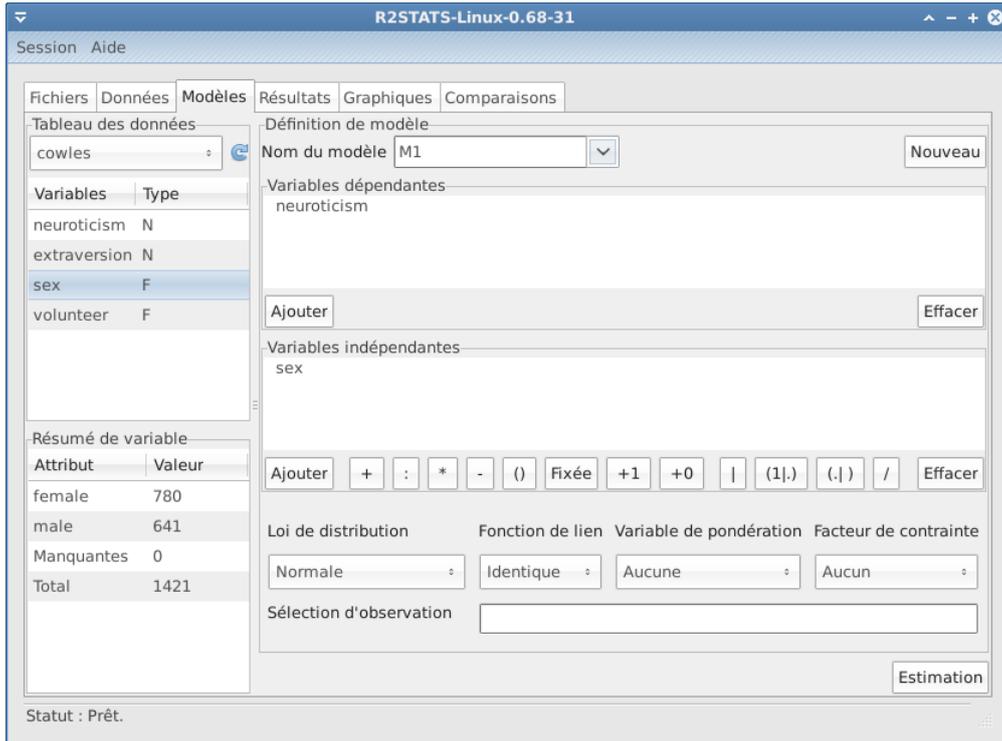


Figure 1: Définition de GLM sous R2STATS

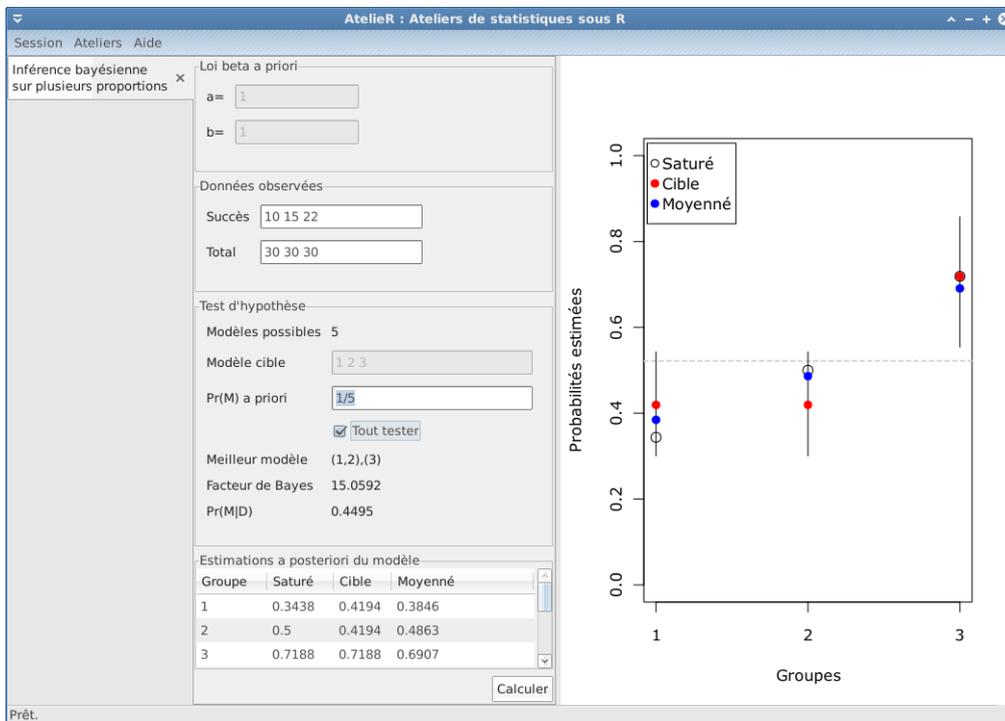


Figure 2: Comparaison bayésienne de proportions binomiales