

Créer un outil pour les non-professionnels de la statistique : Récit d'une expérience

Stéphane HERAULT
contact@adscience.fr

I. INTRODUCTION

Docteur en Biologie-Santé, les quelques années pendant lesquelles j'ai travaillé dans le secteur de la recherche en milieux hospitalier, universitaire et pharmaceutique (à Paris et en province) m'ont donné l'occasion d'apprécier comment les médecins, chercheurs et scientifiques que j'ai côtoyés, utilisent l'outil statistique. La remarque principale est que la plupart des logiciels de statistiques, étant tous des instruments de calcul particulièrement performants et très complets, ne sont par nature accessibles qu'aux professionnels de la statistique ; ce que ne souhaitent pas nécessairement devenir les acteurs de la recherche en France. J'ai listé ci-après les principaux problèmes auxquels ces utilisateurs non-professionnels de la statistique sont confrontés.

Il faut toutefois préciser que, contrairement à bon nombre de mes collègues, j'ai pour ma part "attrapé" le goût de la statistique, grâce aux enseignements des statisticiens du Centre d'Océanologie de Marseille (Station Marine d'Endoume – Université de la Méditerranée).

II. CONSTAT

1. Le choix du test statistique

Lors de ma thèse, ce goût pour la statistique m'a valu ce qui arrive souvent dans un laboratoire de recherche : réaliser l'analyse des données des collègues chercheurs. L'essentiel de mes interventions consistait à les conseiller quant au choix du test statistique le plus approprié, puis à le réaliser et à leur expliquer les résultats. Cette étape du choix du test qui, en théorie, doit être réglée en même temps que la mise au point du protocole, n'est en pratique presque jamais fixée avant que l'utilisateur ne soit confronté au problème. Il m'est également arrivé de recevoir la visite d'un professeur de médecine dont la publication scientifique avait été rejetée par le "reviewer" du journal pour cause de "mauvaise statistique" (i.e. mauvais choix du test statistique utilisé).

Outre un retard des publications, cette habitude de faire l'analyse statistique "comme on peut" ou avec les moyens du bord, peut avoir des conséquences délétères sur la crédibilité qu'engagent le chercheur et son laboratoire vis-à-vis de la communauté scientifique. Par ailleurs, il ne s'agit pas d'une spécificité de la recherche publique puisque je l'ai également rencontrée au sein de laboratoires privés de petite taille, ne disposant pas d'un statisticien à demeure, ni des moyens nécessaires pour s'offrir les services d'une entreprise spécialisée dans l'analyse des données (type CRO : Contract Research Organization).

2. La vérification des conditions d'utilisation

Par ailleurs, dans le cas des tests paramétriques (comparaisons de moyennes, régression linéaire...), la vérification des conditions d'utilisation des tests est une étape ignorée de façon quasi systématique ; non pas par paresse intellectuelle, mais tout simplement parce que beaucoup d'utilisateurs ignorent quelles sont ces conditions (quand ils savent qu'il existe des conditions à respecter). La conséquence est que cette phase est très souvent omise, au risque d'aboutir à une conclusion erronée.