

Quel contenu définir pour un enseignement de statistique destiné à des utilisateurs ? En particulier quelles méthodes présenter, quel niveau de formalisme adopter, quelles justifications théoriques donner ?

Il n'est pas de réponse unique à cette question qui doit être considérée en fonction des utilisateurs (à quel types d'applications seront-ils confrontés ? Quel est leur niveau d'études ?), du temps disponible (deux journées ? trois années ?) et, il faut bien le dire, du formateur.

Ce livre est la transcription du cours de statistique que suivent tous les étudiants d'Agro-campus Rennes. Le contenu a été défini ainsi : la finalité appliquée imposait d'inclure dans l'évaluation des étudiants deux études de cas, l'une à mi-parcours l'autre en fin, menées par les étudiants sur des données qu'ils vont chercher dans les différents laboratoires du campus (dans certains cas, ils réalisent eux-mêmes une petite expérimentation). Ce principe étant acquis, le cours doit les préparer à ces études de cas en présentant :

- un large éventail de méthodes ayant un grand potentiel d'applications ; bien que les enseignants de statistique gardent le contrôle des études de cas choisies, ils n'en ont pas l'initiative ; des situations diverses peuvent apparaître et il faut éviter que l'étudiant s'entende plusieurs fois dire « vous n'avez pas vu les méthodes nécessaires pour traiter ce problème » auquel cas le cours perdrait de sa crédibilité.
- donner une compréhension suffisamment profonde des méthodes pour que l'étudiant puisse reconnaître la problématique d'une méthode dans un questionnement d'utilisateur et bien évaluer la portée des résultats qu'il obtient.

Le volume horaire étant fixé à 18 heures de cours et 38 heures de TD, un compromis drastique a dû être réalisé entre « large éventail » et « compréhension profonde ». Ainsi, presque toutes les démonstrations ont été exclues ainsi que les aspects formels généraux concernant l'estimation et les tests. En contrepartie, le champ couvert comporte, outre les méthodes de statistiques classiques, une ouverture sur les plans d'expériences et l'analyse des données multidimensionnelles.

Le résultat est le présent livre qui est un livre d'apprentissage (le contenu est à la fois centré sur l'essentiel et comporte quelques répétitions - comme dans un cours) et non un livre de référence (tendant vers l'exhaustivité dans un domaine défini).

Puisse ce livre aider le débutant à découvrir la statistique appliquée et, pourquoi pas, à prendre du plaisir dans l'analyse de données.