Principes de la modélisation des données de type temps-évènements : applications en agriculture [Time-to-Event modeling : applications in agriculture]

Gbemavo D.S.JC., Maitre-Assistant de Biométrie, Foresterie (CAMES),
Faculté des Sciences et Techniques (FAST), de l'Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques (UNSTIM), de Dassa Zoumè (Bénin)
Laboratoire de Biomathématiques et d'Estimations Forestières (LABEF) de la Faculté des Sciences Agronomiques / Université d'Abomey-Calavi (Bénin)

Résumé

La présente communication scientifique porte sur l'utilisation d'une méthode statistique couramment utilisée en sciences médicales et dans l'industrie pour traiter les données de type "temps-évènement" mais peu utilisée dans les sciences agronomiques. Les données de type "temps-évènement" sont des données qui expriment le temps (durée) à la survenue d'un évènement (mort, panne, défectuosité, accident, etc.). Ces données sont particulières du fait que la variable d'intérêt n'est pas seulement de savoir si l'évènement a eu lieu ou pas mais aussi quand (temps) l'évènement a eu lieu. La particularité de la méthode présentée est qu'elle tient compte des individus qui n'ont pas présenté "l'évènement" à la fin de l'essai. On parle de données censurées. Ces types de données "temps-évènement" avec censure existent aussi en sciences agronomiques (e.g. temps à : la germination, la mise-bas, la levée, la mort, l'apparition des fleurs, maturation, etc.). En effet, pour la germination, des graines non germées restent généralement à la fin des expériences, et il n'y a aucun moyen d'être certain quand ces graines auraient germé si l'expérience avait été poursuivie indéfiniment. Ces graines sont dites censurées à droite. La présence de telles observations censurées à droite viole les hypothèses de nombreuses méthodes statistiques classiques (modèles linéaires et linéaires généralisés) ce qui fait appel à la méthode d'analyse de survie utilisée pour les données censurées surtout en médecine. Dans l'industrie, on parle d'analyse de « la durée de défection ». L'analyse de survie est caractérisée par deux principales fonctions mathématiques à savoir la fonction de survie et la fonction de risque. La fonction de survie peut être comparée entre deux groupes ou plus à l'aide de statistiques de test basées sur des distributions appropriées. Les effets des covariables discrètes ou continues sur les temps de survie peuvent être quantifiés avec deux types de modèles, le modèle de temps de défaillance accéléré et le modèle de risques proportionnels. La présente communication fera un résumé du principe de l'analyse de survie, et illustrera les aspects ci-dessus cités à partir d'exemples en écologie, production végétale, et zootechnie. Le logiciel statistique R sera utilisé pour les illustrations.

Mots clés : analyse de survie, agriculture, modèles, analyse de données.