



Estimation et tests non paramétriques



Présentation

Description

Les méthodes non-paramétriques sont importantes dans de nombreuses applications statistiques car elles permettent de s'affranchir des approches classiques qui demandent la spécification de modèles statistiques valides. Or établir la validité d'un tel modèle est une entreprise complexe.

Les méthodes non-paramétriques contournent ce problème en utilisant la transformation des données en rang et en conditionnant sur certains quantités issues de la configuration observée de ces rangs. Les statistiques ainsi construites sont indépendantes de la loi des données brutes ce qui permet de construire des procédures d'inférence statistique libre du modèle sous-jacent aux données. En outre, la perte d'efficacité statistique est minimale.

Ce cours constitue une présentation assez complète des méthodes non-paramétriques. Il s'inscrit dans le prolongement d'un premier cours d'introduction aux méthodes inférentielles paramétriques, en adaptant et développant la théorie de plusieurs concepts avancés comme les tests conditionnels, la puissance comparée des tests (mesures d'efficacité), la notion de « effect size ». Il met l'accent sur l'application pratique de ces méthodologies en

faisant un tour d'horizon des principales commandes R et de leurs utilisations.

Objectifs

Amener l'étudiante ou l'étudiant à comprendre les limitations des méthodes statistiques paramétriques, à choisir face à un problème donné une bonne approche non-paramétrique en s'appuyant sur les principes statistiques sous-jacents, à mettre en œuvre via le logiciel R la solution de son problème et à rapporter aux utilisateurs finaux les conclusions de ses analyses et leur portée.

Pré-requis nécessaires

Un cours de statistique inférentielle de niveau M

Un cours de probabilité de niveau L

Pré-requis recommandés : Un cours de statistique descriptive de niveau L

Contrôle des connaissances



Contrôle continu intégral

Syllabus

1. Définitions de *Statistique non paramétrique*

2. Deux astuces pour enlever la dépendance sur un paramètre inconnu : le conditionnement (avec application aux tables de contingence, tests du chi-deux et tests de Fisher-Yates) et la transformation en rang.

3. Le test de Wilcoxon-Mann-Whitney pour 2 échantillons : Les hypothèses et la statistique du test ; Comportement exact et asymptotique sous H_0 , le cas de données ex-æquo, robustesse du test.

4. Variantes et extensions du test de Wilcoxon-Mann-Whitney : Estimation ponctuelle et par intervalle de l'effet des traitements, le cas d'échantillons appariés, puissance du test de Wilcoxon-Mann-Whitney i) quand le paradigme de Student tient et ii) ne tient pas. Notions de « effect-size », calcul de tailles d'échantillons pour obtenir une puissance visée.

5. Autres tests pour le cas de 2 échantillons : test de Kolmogorov-Smirnov, test de Ansari-Bradley

6. Le cas de $K > 2$ échantillons : test de Kruskal-Wallis, test de Friedman-Tukey. Le problème des comparaisons multiples. Contrôle de la p-value : méthode de Holm, méthode FDR (false discovery rate)

7. Indépendance, corrélation et régression : Coefficient de corrélation de Pearson, Spearman, Kruskal, test d'indépendance. Application à la régression

Mise en œuvre avec le logiciel R des principaux tests non-paramétriques vus dans le cours.

Informations complémentaires

Volumes horaires :

CM : 15h

TD : 15h

TP :

Terrain :

Infos pratiques

Contacts

Responsable pédagogique

Elodie Brunel-piccinini

☎ +33 4 67 14 41 64

✉ elodie.brunel-piccinini@umontpellier.fr

Responsable pédagogique

Xavier Bry

☎ +33 4 67 14 35 78

✉ xavier.bry@umontpellier.fr