

Analyse statistique de données

Responsable : Y. Deville (ydeville@ast.obs-mip.fr, 05 61 33 28 24)

Objectifs

Cette unité d'enseignement vise à apporter aux étudiants la maîtrise des outils classiques d'analyse de données expérimentales, ces données pouvant être issues de systèmes extrêmement variés. Sous leur forme la plus simple, les informations ainsi extraites des données rejoignent les résultats d'enquêtes, telles que celles de l'INSEE, auxquels les étudiants peuvent être confrontés quotidiennement : étude de la tendance moyenne d'un paramètre (salaire, âge...) mesuré sur une « population », histogramme des valeurs de ce paramètre... Des outils plus élaborés (tels que l'analyse en composantes principales) sont ensuite considérés, afin de réaliser une analyse plus fine de ces données.

Un autre objectif de cette unité est de présenter aux étudiants la notion de *test d'hypothèse*. Une hypothèse scientifique peut être acceptée ou rejetée en fonction des résultats observés. En statistique, les hypothèses sont décrites par des modèles probabilistes et les données observées sont considérées comme des réalisations des variables aléatoires. L'objet d'un test statistique est de formuler un jugement sur une hypothèse et de distinguer ce qui est plausible de ce qui est peu vraisemblable. Les principaux tests statistiques seront présentés en s'appuyant sur des exemples d'application.

Contenu (C/TD : 16h/10h – TP : 9h)

Cette unité comporte tout d'abord une introduction générale au domaine de la Statistique.

Les outils d'analyse de données expérimentales sont ensuite présentés en plusieurs étapes. On introduit d'abord les paramètres permettant d'analyser les valeurs d'une *unique* variable (moyenne, écart-type, histogramme...). Puis on présente les outils servant à étudier les liaisons entre *deux* variables (analyse de corrélation). Enfin, on en vient au cadre général de l'extraction d'information à partir d'un nombre *quelconque* de variables, en particulier à l'aide de l'analyse en composantes principales.

Un autre domaine relevant du champ général de la Statistique concerne la classification automatique de divers types « d'objets » à partir de paramètres mesurés sur chacun d'eux. Les notions de classification (et de régression) les plus basiques sont présentées dans cette unité. Des approfondissements relatifs à ce domaine seront ensuite fournis dans certains parcours de cette année de Master 2.

La deuxième partie de cette unité commence par un rappel sur les variables aléatoires et les principales lois de probabilité. Ce rappel est suivi par une introduction aux notions générales utilisées dans les tests statistiques telles que *modèles probabilistes, hypothèses nulle et alternative, région de rejet, règles de décision*. On présente ensuite des exemples des tests non-paramétriques, des tests sur des échantillons gaussiens ainsi que des tests paramétriques mettant en compétition deux modèles.

Les divers aspects de la Statistique ainsi présentés conduisent à des applications extrêmement variées. Différentes applications de ce type sont présentées dans les travaux dirigés et travaux pratiques de cette unité.

Pré-requis

Cette UE n'a pas de pré-requis spécifiques.

Bibliographie

G. Saporta , « Probabilités, analyse des données et statistique », Technip, Paris, 1990.

E. Lehmann and J. P. Romano, « Testing statistical hypotheses », 3rd ed., New York: Springer, 2005.