

Econométrie des variables qualitatives

Infos pratiques

- > ECTS : 3.0
- > Nombre d'heures : 24.0
- > Langue(s) d'enseignement : Français
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral
- > Ouvert aux étudiants en échange : Oui
- > Composante : Sciences économiques, gestion, mathématiques et informatique

Présentation

- Séance # 1 : Qu'est-ce que la modélisation de données ? Les différents types de données.
 - Le questionnement statistique à travers quelques exemples de la vie réelle (en économie, en sociologie, finance, etc avec de vraies données)
 - Introduction au logiciel R et à l'environnement RStudio pour comprendre des demos, résultats et quelques séances de TD qui seront proposés dans ce cours.
- Séance # 2 : Visualisation et analyse de données (différents types de variables : quantitatives et qualitatives),
 - Construction de sous échantillons, manipulation de tableaux, résumés, stat descriptives avec de demos en directe sur le logiciel R.
 - Visualisation : diagrammes en bâtons, nuage de points, boîte à moustaches, nuage de points, nuage de mots, etc.
 - Analyse de données : Comparer des groupes (visuellement avec des graphiques pertinents)
- Séance # 3 : Analyse multivariée et Clustering
 - Visualisation, Projections
 - ACP
 - Clustering (Kmeans,...)
- Séance # 4 : Test d'hypothèses.
 - Qu'est-ce que la p-valeur ? (Motivation avec des simulations sur R)
 - Test de comparaison de moyennes et test de comparaison proportions avec des données directement sur R.
 - Test d'Indépendance pour deux variables qualitatives et quantitatives. Motivation toujours avec des données socio-économiques et des résultats sur le logiciel R.
- Séance # 5 : Techniques de Re-échantillonnage et validité des résultats : inférences classiques, (bootstrap, validation croisée).
- Séance # 6 : Régression linéaire simple et multiple.
 - Présentation du modèle et de ses objectifs.
 - Prédiction.
 - Estimation de l'erreur de prédiction à l'aide des techniques de ré-échantillonnage
- Séance # 7 : Choix du modèle
 - Sélection d'un bon modèle via la métrique de l'estimation de l'erreur.
 - Compétition entre plusieurs modèles dans la famille de régression d'un point de vue plutôt pratique avec sorties du logiciel R.
- Séance # 8: Régression logistique (logit).
 - Présentation du modèle et
 - Interprétation du rapport de chances (Odds-ratio)
 - Sensibilité et spécificité d'un test diagnostique : interprétation des taux de faux négatifs et faux positifs...

Objectifs

En statistiques on se pose des questions sur les données, et on essaie d'y répondre. L'objectif de ce cours est d'introduire les principaux éléments d'analyse de données et d'illustrer quelques concepts et outils

fondamentaux de la statistique et la modélisation.
L'accent est mis sur la pratique plus que sur la théorie statistique-mathématique. L'objectif est que les étudiants sachent utiliser les modèles statistiques classiques (tests d'hypothèses, régressions, modèle logit, probit, sélection de modèles,...) avec des applications sur des données socio-economiques et autres.

Évaluation

Examen écrit : 100%

Pré-requis nécessaires

Ce cours peut être suivi par des étudiants n'ayant qu'une connaissance basique des statistiques (au moins les concepts de population, échantillon, proportions, moyennes et variances, représentation graphiques, et quelques distributions connues en stat ...) grâce à des synthèses présentées en début de séances.

Compétences visées

- Comprendre le vocabulaire et les concepts fondamentaux de l'analyse de données.
- Se poser des questions sur les données.
- Analyser les données d'un point de vue statistique, modéliser, interpréter et répondre aux questions posées, expliquer les résultats obtenus par le logiciel R, Rstudio.

Bibliographie

- Jean-Herman Guay. *Statistiques en sciences humaines avec R (sciences sociales). 2eme édition*
- Pierre-André Cornillon, Eric Matzner-Lober. *Régression (Théorie et applications).*
- <http://factominer.free.fr/graphs/factoshiny.html>
- J-H Guay (voir le site <http://dimension.usherbrooke.ca/dimension/v2ssrcadre.html>)

Ressources pédagogiques

- Diapositives des rappels et du cours
- Mises en œuvre corrigées avec R pour quelques séances de TD