

Réaliser ses analyses statistiques avec R

(3 jours)

Claire Della Vedova (claire@delladata.fr) Data Value (formation@datavalue.fr)

Objectifs

- Apprendre à utiliser le logiciel R pour analyser des données.
- Mettre en oeuvre dans R les méthodes de statistique descriptive, décisionnelle, analyse de la variance, régression linéaire et analyse de données multidimensionnelles.

Public

Toute personne souhaitant analyser des données avec R

Prérequis

Pour suivre ce stage dans de bonnes conditions, il est recommandé d'avoir suivi en amont la formation R – [Prise en main, analyses statistiques et graphiques]<https://www.datavalue.fr/formation-r-prise-en-main-analyses-statistiques-et-graphiques>

Méthode

Pédagogie active mêlant exposés, exercices et applications pratiques dans le logiciel R.

Modalités d'évaluation

Un questionnaire d'auto-évaluation proposé en amont de la formation nous permettra d'évaluer votre niveau et de recueillir vos attentes. Ce même questionnaire soumis en aval de la formation nous permettra d'apprécier votre progression.

Accessibilité

Vous souhaitez suivre notre formation Réaliser ses analyses statistiques avec R et êtes en situation de handicap ? Merci de nous contacter afin que nous puissions envisager les adaptations nécessaires et vous garantir de bonnes conditions d'apprentissage

Programme (3jours, soit 21h)

Statistiques descriptives

- Gestion d'un jeu de données - dataframe : Importation, caractérisation, sélection, sous-ensembles
- Premières analyses d'un jeu de données : Premières vérifications, valeurs manquantes, recodage
- Résumé d'une variable quantitative – numeric : Indicateurs numériques, représentations graphiques
- Résumé d'une variable qualitative – factor Indicateurs numériques, représentations graphiques

Intervalle de confiance

- Le raisonnement à partir d'un échantillon : Généralités, échantillonnage, estimation d'un paramètre
- Intervalle de confiance d'une moyenne
- Intervalle de confiance d'une proportion
- Intervalle de confiance d'une variance

Tests d'hypothèses

- Qu'est-ce qu'un test d'hypothèse ? : Généralités, règle de décision, risques d'erreur, puissance
- Les tests de conformité ou de comparaison à une norme : Conformité d'une moyenne, d'une proportion
- Les tests de comparaison de deux populations : Comparaison de deux moyennes, de deux proportions
- Test d'ajustement à une loi de probabilité normale : Le test de Shapiro-Wilk
- Introduction aux tests non paramétriques : Test de Wilcoxon, Mann et Whitney, Kruskal Wallis, Friedman

Liaisons entre deux variables

- Liaison entre deux variables quantitatives : Nuage de points, intensité de la liaison, significativité
- Liaison entre deux variables qualitatives : Tableau de contingence, intensité et significativité du lien de dépendance : test du khi2
- Liaison entre une variable qualitative et quantitative : Comparaison de plusieurs populations, le rapport de corrélation
- Liaisons entre plusieurs variables - Approches graphiques : matrice de nuages de points, treillis - Caractériser des sous-populations par plusieurs variables

L'analyse de la variance – Anova

- Analyse de la variance à un facteur : Variabilité inter, intra, totale – Rapport de corrélation - Le test de Fisher
- Comparaisons multiples de moyennes : Analyses post hoc, la procédure de Tukey
- Analyse de la variance à deux facteurs et interaction
- Extensions de l'Anova : Modèle à effet fixe ou aléatoire, modèle hiérarchisé

Régression linéaire simple et multiple

- De la corrélation à la régression : L'intérêt d'un modèle - Variables à expliquer, explicatives, erreur
- La régression linéaire simple : Ajustement par la méthode des moindres carrés - Tests et validation du modèle
- La régression linéaire multiple

- Choix d'un modèle de régression : Pourquoi sélectionner un sous-ensemble de variables explicatives ?
Les différentes approches et critères de sélection d'un modèle

Analyse de données multidimensionnelles

- Un panorama des méthodes multidimensionnelles Analyses factorielles, classification – Le package FactoMineR
- ACP : Analyse en Composantes principales
- AFC : Analyse Factorielle des Correspondances
- ACM : Analyse des correspondances Multiples
- CAH : Classification Ascendante Hiérarchique