

Arbres de décision pour la découverte de profils vulnérables

Emmanuel Rousseaux

Institut d'études démographiques et du parcours de vie
Université de Genève
1211 Genève 4, Suisse
emmanuel.rousseaux@unige.ch

Plan de présentation

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

NCCR LIVES : « Surmonter la vulnérabilité : perspective du parcours de vie »

Travaille au sein de l'IP « méthodes » dirigée par Gilbert Ritschard

Vise à fournir des méthodes d'analyse du parcours de vie pour l'identification des dynamiques des vulnérabilités

NCCR LIVES : « Surmonter la vulnérabilité : perspective du parcours de vie »

Travaille au sein de l'IP « méthodes » dirigée par Gilbert Ritschard

Vise à fournir des méthodes d'analyse du parcours de vie pour l'identification des dynamiques des vulnérabilités

NCCR LIVES : « Surmonter la vulnérabilité : perspective du parcours de vie »

Travaille au sein de l'IP « méthodes » dirigée par Gilbert Ritschard

Vise à fournir des méthodes d'analyse du parcours de vie pour l'identification des dynamiques des vulnérabilités

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les prochaines années
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les prochaines années
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les prochaines années
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ **Vulnérabilité : concept à opérationnaliser**
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les p années à venir
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les p années à venir
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les p années à venir
- ▶ ...

Découvrir quels sont les profils de personnes vulnérables, comment sont-elles tombées dans la vulnérabilité, et comment certaines de ces personnes ont réussies à s'en sortir.

- ▶ Vulnérabilité : concept à opérationnaliser
- ▶ Personnes en dessous d'un certain revenu, état de santé, satisfaction, ...
- ▶ Probabilité de subir tel événement non désirable dans les p années à venir
- ▶ ...

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ Fouilles multicanale du parcours de vie
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Développer des outils de visualisation

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ Fouilles multicanale du parcours de vie
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Développer des outils

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ Fouilles multicanale du parcours de vie
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Découvrir de nouvelles hypothèses à tester

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ **Fouilles multicanale du parcours de vie**
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Découvrir de nouvelles hypothèses à tester

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ Fouilles multicanale du parcours de vie
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Découvrir de nouvelles hypothèses à tester

Méthodologie générale de fouille de données

Axes principaux du projet

- ▶ Étendre les méthodes d'analyses de séquences aux variables quantitatives
- ▶ Fouilles multicanale du parcours de vie
- ▶ Faciliter l'interopérabilité entre jeux de données
- ▶ Découvrir de nouvelles hypothèses à tester

Motivation : découvrir des (les) profils de personnes subissant une vulnérabilité donnée

En pratique

- Découvrir des sous groupes de la population
- où la proportion de personnes vulnérables est significativement plus importante que dans la population générale
- à partir de données de type "questionnaire"

Motivation : découvrir des (les) profils de personnes subissant une vulnérabilité donnée

En pratique

- Découvrir des sous groupes de la population
- où la proportion de personnes vulnérables est significativement plus élevée que dans la population générale
- à l'aide d'arbres de décision

Motivation : découvrir des (les) profils de personnes subissant une vulnérabilité donnée

En pratique

- ▶ Découvrir des sous groupes de la population
- ▶ où la proportion de personnes vulnérables est significativement plus importante que dans la population générale
- ▶ et (si possible) où des profils proches ont évité la vulnérabilité

Motivation : découvrir des (les) profils de personnes subissant une vulnérabilité donnée

En pratique

- ▶ Découvrir des sous groupes de la population
- ▶ où la proportion de personnes vulnérables est significativement plus importante que dans la population générale
- ▶ et (si possible) où des profils proches ont évité la vulnérabilité

Motivation : découvrir des (les) profils de personnes subissant une vulnérabilité donnée

En pratique

- ▶ Découvrir des sous groupes de la population
- ▶ où la proportion de personnes vulnérables est significativement plus importante que dans la population générale
- ▶ et (si possible) où des profils proches ont évité la vulnérabilité

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

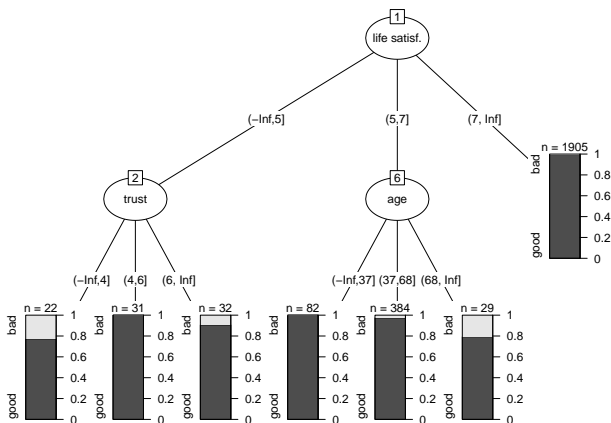


Figure: Profils de santé. Données du PSM 2011, vague de 2010

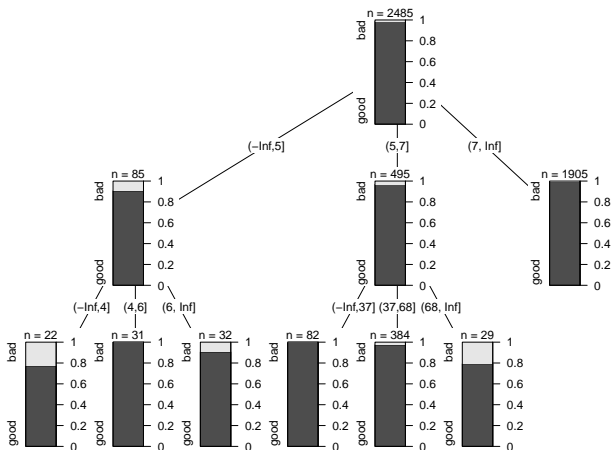


Figure: Profils de santé. Données du PSM 2011, vague de 2010

Processus de construction récursif

- ▶ Point de départ (1er nœud) : population entière
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ On retient le descripteur avec la meilleure performance
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

Processus de construction récursif

- ▶ **Point de départ (1er nœud) : population entière**
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ On retient le descripteur avec la meilleure performance
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

Processus de construction récursif

- ▶ Point de départ (1er nœud) : population entière
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ On retient le descripteur avec la meilleure performance
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

Processus de construction récursif

- ▶ Point de départ (1er nœud) : population entière
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ **On retient le descripteur avec la meilleure performance**
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

Processus de construction récursif

- ▶ Point de départ (1er nœud) : population entière
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ On retient le descripteur avec la meilleure performance
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

Processus de construction récursif

- ▶ Point de départ (1er nœud) : population entière
- ▶ Pour chaque descripteur : choix du *meilleur* découpage ou regroupement
- ▶ On retient le descripteur avec la meilleure performance
- ▶ Si aucun critère d'arrêt n'est satisfait, on éclate suivant le descripteur retenu
- ▶ On répète le processus dans chaque nœuds fils.

On évalue généralement la qualité d'un éclatement à l'aide

- ▶ d'une mesure de la qualité d'un nœud
- ▶ et d'une mesure de la qualité du passage du nœud parent à l'éclatement fils

On évalue généralement la qualité d'un éclatement à l'aide

- ▶ d'une mesure de la qualité d'un nœud
- ▶ et d'une mesure de la qualité du passage du nœud parent à l'éclatement fils

On évalue généralement la qualité d'un éclatement à l'aide

- ▶ d'une mesure de la qualité d'un nœud
- ▶ et d'une mesure de la qualité du passage du nœud parent à l'éclatement fils

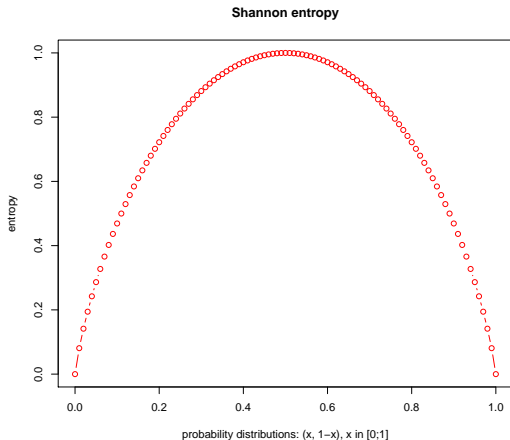


Figure: Entropie de Shannon - cas binaire

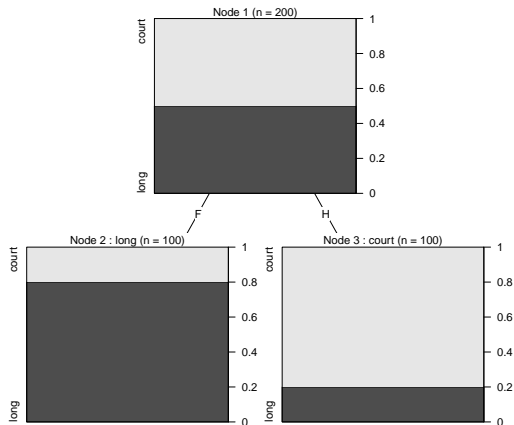


Figure: Prédiction de la longueur des cheveux (données arbitraires)

Pour évaluer la qualité globale de l'éclatement (vis-à-vis du parent), deux principales mesures :

- ▶ Gain d'information
- ▶ Gain ratio

Pour évaluer la qualité globale de l'éclatement (vis-à-vis du parent), deux principales mesures :

- ▶ **Gain d'information**
- ▶ Gain ratio

Pour évaluer la qualité globale de l'éclatement (vis-à-vis du parent), deux principales mesures :

- ▶ Gain d'information
- ▶ **Gain ratio**

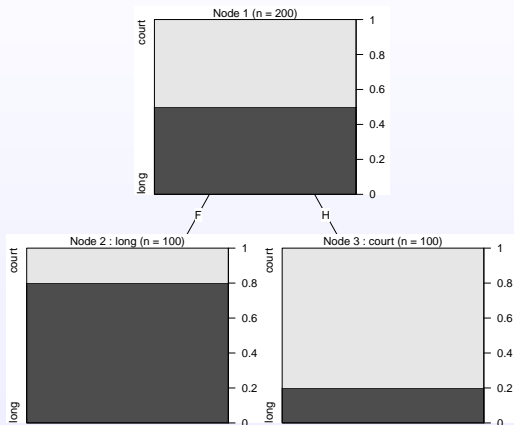


Figure: Entropie de Shannon - x - cas binaire

Tester tous les regroupement en 2, 3, 4, ... des valeurs/modalités du descripteur

- ▶ Cas numérique ou catégoriel ordinal : regroupements adjacents
- ▶ Cas catégoriel nominal : toutes les permutations sont possibles

Tester tous les regroupement en 2, 3, 4, ... des valeurs/modalités du descripteur

- ▶ Cas numérique ou catégoriel ordinal : regroupements adjacents
- ▶ Cas catégoriel nominal : toutes les permutations sont possibles

Lorsqu'une classe est sous représentée dans les données

- ▶ Performances moyennes sur la classe minoritaire
- ▶ Qui est portant notre classe d'intérêt
- ▶ On ne devrait pas essayer de s'éloigner du 50/50 !

⇒ Idée de décentrée les entropies

Lorsqu'une classe est sous représentée dans les données

- ▶ Performances moyennes sur la classe minoritaire
- ▶ Qui est portant notre classe d'intérêt
- ▶ On ne devrait pas essayer de s'éloigner du 50/50 !

⇒ Idée de décentrée les entropies

Lorsqu'une classe est sous représentée dans les données

- ▶ Performances moyennes sur la classe minoritaire
- ▶ Qui est portant notre classe d'intérêt
- ▶ On ne devrait pas essayer de s'éloigner du 50/50 !

⇒ Idée de décentrée les entropies

Lorsqu'une classe est sous représentée dans les données

- ▶ Performances moyennes sur la classe minoritaire
- ▶ Qui est portant notre classe d'intérêt
- ▶ On ne devrait pas essayer de s'éloigner du 50/50 !

⇒ Idée de décentrée les entropies

Lorsqu'une classe est sous représentée dans les données

- ▶ Performances moyennes sur la classe minoritaire
- ▶ Qui est portant notre classe d'intérêt
- ▶ On ne devrait pas essayer de s'éloigner du 50/50 !

⇒ Idée de décentrée les entropies

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

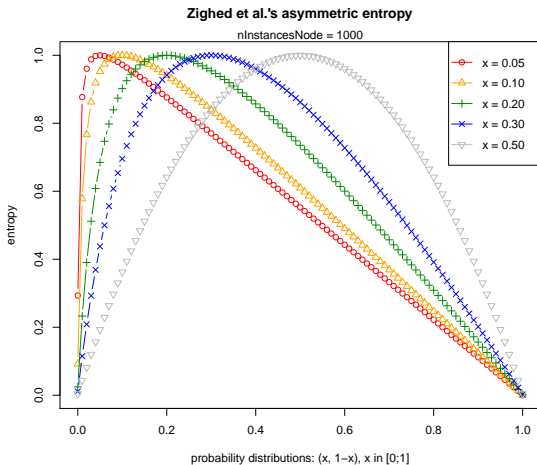


Figure: Entropie assymétrique [Zighed et al., 2007] pour différents déséquilibres

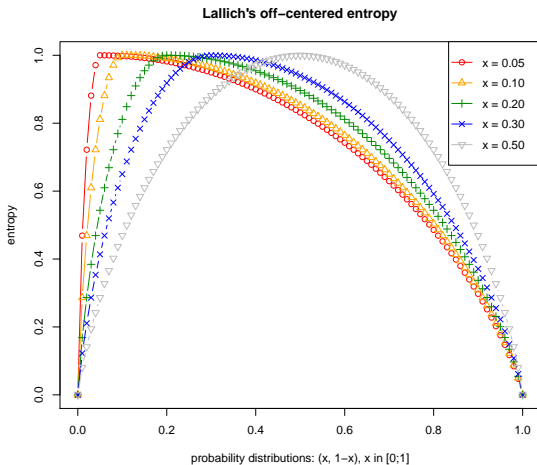


Figure: Entropie décentré [Lenca et al., 2010] pour différents déséquilibres

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ **Nombre minimal d'individu par feuille à adapter ?**
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

Les expérimentations des auteurs ont montrées une amélioration des résultats, mais on pourrais espérer mieux

Raisons formulées

- ▶ Effet résiduel du décentrage ?
- ▶ Effet résiduel de la consistance ?
- ▶ Nombre minimal d'individu par feuille à adapter ?
- ▶ Cas binaire symétrique, méthodes plus adaptés au multi-classe

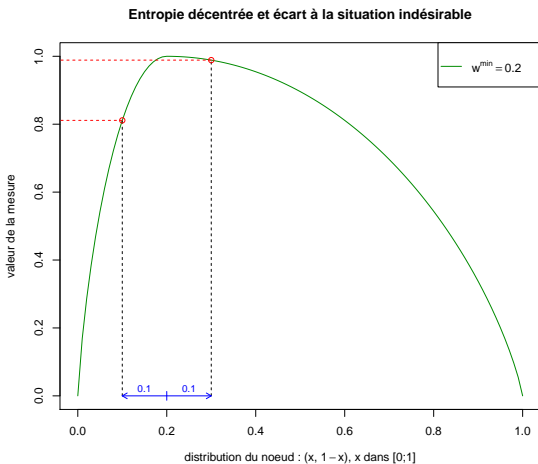


Figure: Effet résiduel du décentrage

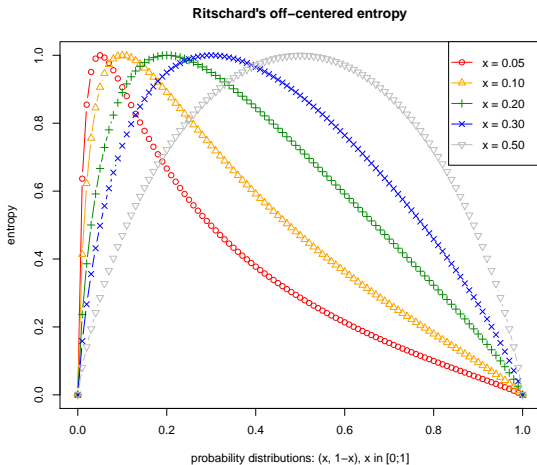


Figure: Entropie décentrée par changement de variable

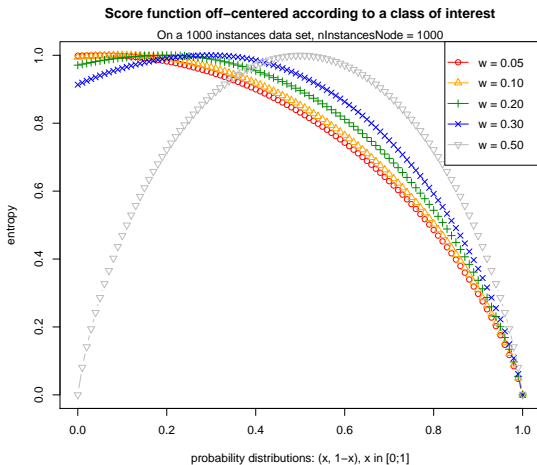


Figure: *Score décentré*

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

À ce jour, aucune implémentation libre des méthodes décentrées

- ▶ Proposer les méthodes à la communauté scientifique
- ▶ Tester les nouvelles propositions
- ▶ Mieux comprendre les spécificités de chaque méthode

À ce jour, aucune implémentation libre des méthodes décentrées

- ▶ Proposer les méthodes à la communauté scientifique
- ▶ **Tester les nouvelles propositions**
- ▶ Mieux comprendre les spécificités de chaque méthode

À ce jour, aucune implémentation libre des méthodes décentrées

- ▶ Proposer les méthodes à la communauté scientifique
- ▶ Tester les nouvelles propositions
- ▶ Mieux comprendre les spécificités de chaque méthode

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ **Toutes les méthodes discutées ici sont présentes**
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ Validation croisée
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ Validation croisée
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ **Éclatements n -aires**
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ Validation croisée
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ Validation croisée
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ **Indicateurs de qualité de l'arbre**
- ▶ Validation croisée
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ **Validation croisée**
- ▶ ...

Les méthodes sont disponibles sous la forme d'une librairie R

- ▶ Toutes les méthodes discutées ici sont présentes
- ▶ Possibilité de définir soit même l'indicateur de qualité
- ▶ Éclatements n -aires
- ▶ Méthodes d'affichage de l'arbre
- ▶ Indicateurs de qualité de l'arbre
- ▶ Validation croisée
- ▶ ...

En cours !

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

- ▶ **Poursuivre les expérimentations**
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ Et surtout :

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

- ▶ Poursuivre les expérimentations
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ Et surtout :

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

- ▶ Poursuivre les expérimentations
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ Et surtout :

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

- ▶ Poursuivre les expérimentations
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ **Et surtout :**

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

- ▶ Poursuivre les expérimentations
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ Et surtout :

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

- ▶ Poursuivre les expérimentations
- ▶ Publier les résultats des comparaisons des méthodes
- ▶ Diffuser la librairie sur le CRAN
- ▶ Et surtout :

Appliquer ces méthodes dans le cadre de LIVES pour la détection des profils vulnérables (en transversal et longitudinal)

Plan

Introduction

Arbres de décision

Entropies décentrées

La librairie Trim

Perspectives

Bonus

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ **Faire adopter R !**
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ **Fonctions orientées utilisateur**
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ **Vus synthétiques des données**
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ **Tests automatiques de consistance**
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ Documentations et guide de démarrage détaillé

Motivation : fournir un cadre de manipulation des données adéquat pour le sociologue

- ▶ Faire adopter R !
- ▶ Gestion plus sophistiquée des valeurs manquantes
- ▶ Fonctions orientées utilisateur
- ▶ Vus synthétiques des données
- ▶ Tests automatiques de consistance
- ▶ Tests automatiques de représentativité des données
- ▶ **Documentations et guide de démarrage détaillé**

Dataset : cadre de manipulation de données transversales

- ▶ Manipulation un peu à la SPSS
- ▶ Types de valeurs manquantes personnalisables
- ▶ Export en pdf des vues synthétiques
- ▶ Gestion automatique des poids
- ▶ ...

Dataset : cadre de manipulation de données transversales

- ▶ **Manipulation un peu à la SPSS**
- ▶ Types de valeurs manquantes personnalisables
- ▶ Export en pdf des vues synthétiques
- ▶ Gestion automatique des poids
- ▶ ...

Dataset : cadre de manipulation de données transversales

- ▶ Manipulation un peu à la SPSS
- ▶ **Types de valeurs manquantes personnalisables**
- ▶ Export en pdf des vues synthétiques
- ▶ Gestion automatique des poids
- ▶ ...

Dataset : cadre de manipulation de données transversales

- ▶ Manipulation un peu à la SPSS
- ▶ Types de valeurs manquantes personnalisables
- ▶ **Export en pdf des vues synthétiques**
- ▶ Gestion automatique des poids
- ▶ ...

Dataset : cadre de manipulation de données transversales

- ▶ Manipulation un peu à la SPSS
- ▶ Types de valeurs manquantes personnalisables
- ▶ Export en pdf des vues synthétiques
- ▶ **Gestion automatique des poids**
- ▶ ...

Dataset : cadre de manipulation de données transversales

- ▶ Manipulation un peu à la SPSS
- ▶ Types de valeurs manquantes personnalisables
- ▶ Export en pdf des vues synthétiques
- ▶ Gestion automatique des poids
- ▶ ...

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ Extraction longitudinale
- ▶ Interopérabilité avec les outils d'analyses longitudinales

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ Extraction longitudinale
- ▶ Interopérabilité avec les outils d'analyses longitudinales

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ Extraction longitudinale
- ▶ Interopérabilité avec les outils d'analyses longitudinales

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ Extraction longitudinale
- ▶ Interopérabilité avec les outils d'analyses longitudinales

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ **Extraction longitudinale**
- ▶ Interopérabilité avec les outils d'analyses longitudinales

ℓ Dataset : cadre de manipulation de données longitudinales

- ▶ Basée sur Dataset
- ▶ Visualisation longitudinale du jeu de données
- ▶ Aide de construction d'un sous jeu de données
- ▶ Extraction longitudinale
- ▶ Interopérabilité avec les outils d'analyses longitudinales

Éléments de bibliographie

Éléments de bibliographie I

- [Zighed et al., 2007] Djamel-Abdelkader Zighed and Simon Marcellin and Gilbert Ritschard *Mesure d'entropie asymétrique et consistante*. 7ème Conférence Extraction et Gestion des Connaissances (EGC'07) :81–86 Revue des Nouvelles Technologies de l'Information.(2007). Cépaduès.
- [Ritschard et al., 2007] Gilbert Ritschard and Djamel-Abdelkader Zighed and Simon Marcellin *Données déséquilibrées, entropie décentrée et indice d'implication*. Nouveaux apports théoriques à l'analyse statistique implicative et applications :315–327 (2007). ASI4. Departament de Matemàtiques, Universitat Jaume I.
- [Marcellin, 2008] Simon Marcellin. *Arbres de décision en situation d'asymétrie*. Université Lumière Lyon II (2008).
- [Ritschard, 2008] Gilbert Ritschard and Vincent Pisetta and Djamel-Abdelkader Zighed. *Inducing and evaluating classification trees with statistical implicative criteria*. Statistical Implicative Analysis : Theory and Applications, 127/2008 :397–419. (2008). Studies in Computational Intelligence, Springer.
- [Zighed et al.,2009] Djamel A. Zighed, Gilbert Ritschard, Simon Marcellin. *Asymmetric and sample size sensitive entropy measures for supervised learning*. Studies in Computational Intelligence (2009). Springer

Éléments de bibliographie II

- [Ritschard et al., 2009] Gilbert Ritschard and Simon Marcellin and Djamel-Abdelkader Zighed *Arbre de décision pour données déséquilibrées : sur la complémentarité de l'intensité d'implication et de l'entropie décentrée*. *Revue des Nouvelles Technologies de l'Information*, E :213–226. (2009). Cépaduès.
- [Lenca et al., 2010] Philippe Lenca, Stéphane Lallich, B Vaillant *Construction of an off-centered entropy for the supervised learning of imbalanced classes : Some first results*. *Communications in Statistics - Theory and Methods* (2010). Editeur.
- [Losa et al., 2005] Fabio Losa and Pau Origoni and Gilbert Ritschard. *Arbres d'induction dans un contexte non classificatoire*. EGC, janvier 2005.
- [Hothorn et al., 2009] Torsten Hothorn and Achim Zeileis. *partykit : A Toolbox for Recursive Partytionning* (2009)
<http://www.agrocampus-ouest.fr/math/user-2009/slides/Hothorn+Zeileis.pdf>.

Merci de votre attention

Des questions ?