



7·8·9
NOVEMBRE
2022

42es Journées Annuelles
de la Société Française
De Gériatrie et de Gérontologie
Journées de formation et de recherche en gériatrie
Centre de Conférence : Hyatt Regency Paris Étoile

EN SAVOIR + ?

S'INSCRIRE

Quels modèles statistiques simples à utiliser ?

Recherche Clinique en Gériatrie

JB. Beuscart – CHU de Lille – Université de Lille

jean-baptiste.beuscart@univ-lille.fr

Les statistiques dans la recherche médicale

Méthodes statistiques multivariées

Michaël Genin, Alain Duhamel, Patrick Devos

Université de Lille 2

EA 2694 - Santé Publique : Epidémiologie et Qualité des soins

michael.genin@univ-lille2.fr



Michaël GENIN

Maître de conférences en statistique - Associate Professor of Statistics

CNU : Section 26 - Mathématiques appliquées et applications des mathématiques



LABORATOIRE(S) / EQUIPE(S)
[i] Equipe d'accueil 2694 - Santé publique : épidémiologie et qualité des soins
COORDINATION(S) DE FORMATION
[i] Faculté de Médecine Henri Warembourg
[i] Université Lille Droit et Santé

+33 (0)3 20 62 68 82
[i] Mail
Faculté de Médecine - Pôle Recherche (Dôme Mag) 1 place de Verdun, F-59045 Lille

DOMAINE(S) DE RECHERCHE

Biostatistics - Spatial statistics - Scan statistics - Spatial epidemiology

PRÉSENTATION

AXES DE RECHERCHE

ENSEIGNEMENTS

PUBLICATIONS

PARCOURS

Vous retrouverez au sein de cette section les supports de cours relatifs aux différentes formations dans lesquelles j'interviens. Pour chaque cours, deux versions sont disponibles : une version avec animations et une version sans animation (version imprimable). Mes supports de cours évoluent d'une année sur l'autre, veuillez bien faire attention à la date de version disponible sur chacun des supports.

Si vous avez des questions, remarques, corrections, n'hésitez pas à m'envoyer un [eMail](#).

Gestion de données

- Conception d'un formulaire papier - [i](#) - [slides](#)
- Consignes pour la saisie de données - [i](#), [slides](#) - [i](#), [data](#)

Statistique descriptive

- Cours de statistique descriptive - [i](#), [slides](#) - [version imprimable](#)

Probabilités - Variables Aléatoires - Loi usuelles

- Quelques éléments pour le calcul des probabilités - [i](#), [slides](#) - [i](#), [version imprimable](#)
- Variables aléatoires - [i](#), [slides](#) - [i](#), [version imprimable](#)
- Loi de probabilité usuelles - [i](#), [slides](#) - [version imprimable](#)

Inférence statistique (Estimation ponctuelle et par intervalle de confiance)

- Théorie de l'estimation - [i](#), [slides](#) - [version imprimable](#)

Tests statistiques

- Tests paramétriques (principes, comparaison de moyennes, comparaison de fréquences...) - [i](#), [slides](#) - [version imprimable](#)
- Tests du KHI-Deux - [i](#), [slides](#) - [i](#), [version imprimable](#)
- Tests non-paramétriques - [i](#), [slides](#) - [i](#), [version imprimable](#)
- TP Tests NP - Sujet - Données - [i](#), Anxiété - [i](#), Gontages - [i](#), Juges

Modélisation

- Corrélation / Régression linéaire simple et multiple - [i](#), [slides](#) - [i](#), [version imprimable](#)
- Régression logistique binaire - [i](#), [slides](#) - [version imprimable](#) - Sujet TP - [i](#), Données TP
- Introduction à l'analyse de survie - [i](#), [slides](#) - [i](#), [version imprimable](#) - Sujet TP - [i](#), Données TP
- Aires de classement - (CHAID) - [i](#), [slides](#) - [version imprimable](#)
- Analyse Factorielle Discriminante - [i](#), [slides](#) - [i](#), [version imprimable](#)

Biostatistiques

- Evaluation d'un test diagnostique - Concordance - [i](#), [slides](#) - [version imprimable](#)
- Introduction aux méthodes statistiques multivariées (M2 Recherche Biologie et Santé) - [i](#), [slides](#) - [i](#), [version imprimable](#)

Data Mining

- Introduction à la fouille de données (Journées thématiques M2 Biologie et Santé) - [i](#), [slides](#) - [i](#), [version imprimable](#)

Plan de la présentation

- Objectif : Présenter une vue générale des modèles statistiques simples à utiliser en recherche clinique en médecine
 1. Enjeux
 2. Variables
 3. Méthodes
 4. Perspectives

Plan de la présentation

- Objectif : Présenter une vue générale des modèles statistiques simples à utiliser en recherche clinique en médecine
 1. Enjeux
 2. Variables
 3. Méthodes
 4. Perspectives

La Statistique et les Biostatistiques

- La STATISTIQUE : discipline traitant du **recueil** (plans d'expérience, sondages, ...), du **traitement** et de **l'interprétation** de données caractérisées par une grande **variabilité**.
- Partie des mathématiques appliquées, utilisant la théorie des probabilités.
- Beaucoup de domaines d'applications
 - Sondages : enquêtes d'opinion
 - Industrie : contrôle de qualité
 - Marketing : scoring, profil de consommateurs
 - **Médecine : épidémiologie, recherche clinique**
 -
- Statistiques appliquées à la Médecine = **BIOSTATISTIQUES**
 - Données spécifiques : variabilité inter et intra, données interprétées, ...
 - Méthodes spécifiques : survie, courbes ROC, plans d'expérience...

Méthodologie statistique

- Employer bien sûr la "bonne" procédure statistique pendant l'analyse !!!

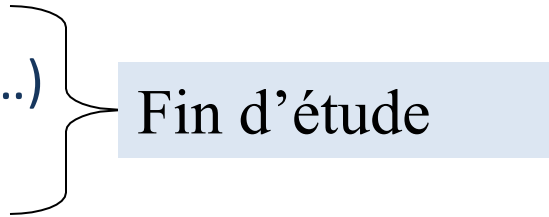
- **MAIS cela ne suffit pas ...**

- Choisir le bon type d'étude
- Choisir le bon plan d'expérience
- Choisir les bons critères de jugement
- Définir les variables recueillies
- Qualité des données recueillies



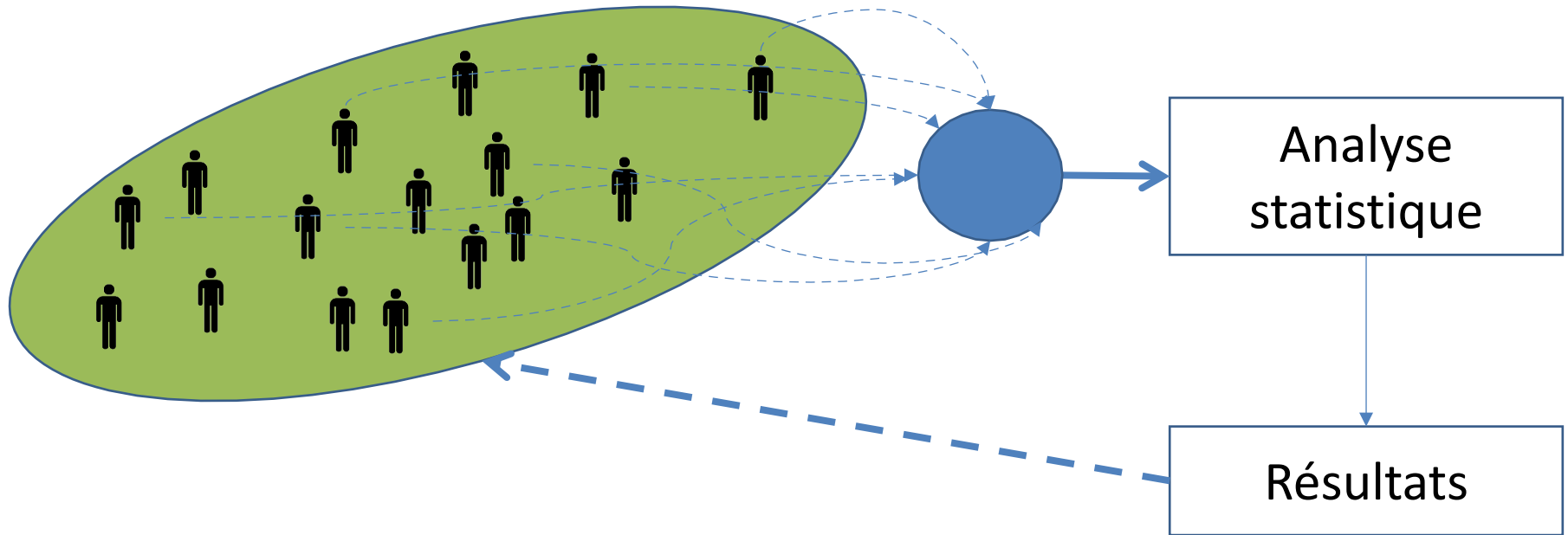
Avant l'étude !!!

- Analyse statistique rigoureuse (tests, modèles, ...)
- Bonne interprétation des résultats



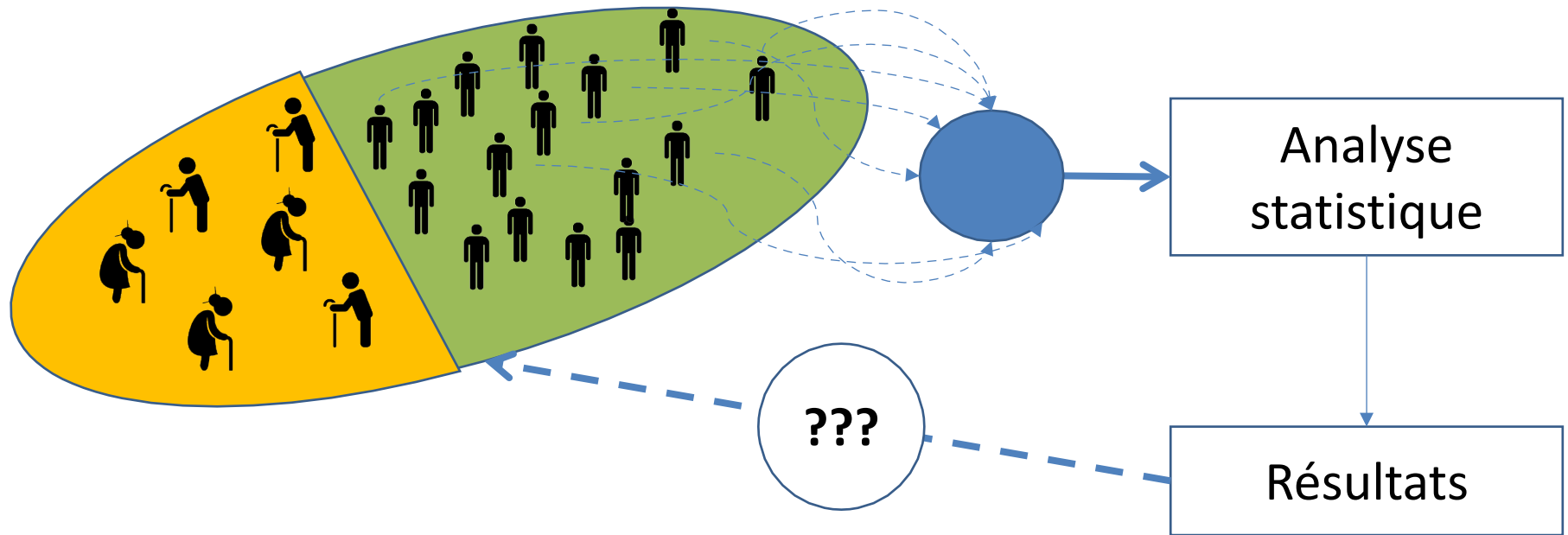
Fin d'étude

Inférence statistique



- ◆ On désire étudier une population P :
 - On tire un échantillon E de taille n issu de P
 - On analyse les caractéristiques de E
 - On généralise à P

Inférence statistique



Définir très précisément la population que l'on désire étudier !!

Définir très précisément les mesure effectuées sur la population étudiée

!!

Messages

- 1) LA base : qualité des méthodes d'échantillonnage et de mesure des données recueillies au cours de l'étude
- 2) Une analyse statistique de haute qualité sur des données de mauvaise qualité = étude de mauvaise qualité

Plan de la présentation

- Objectif : Présenter une vue générale des modèles statistiques simples à utiliser en recherche clinique en médecine
 1. Enjeux
 2. Variables
 3. Méthodes
 4. Perspectives

Méthodes Statistiques : définitions générales

- **INDIVIDU** : « Objet » sur lequel un ou plusieurs caractères peuvent être observés.
- **POPULATION** : Ensemble des individus pris en considération.
- **VARIABLE** : peut être **qualitative** (attribut) ou **quantitative** (numérique).

Différents types de variables

- Variables Quantitatives
 - Variables quantitatives continues (âge, poids, taille,)
 - Variables quantitatives discrètes (ne peuvent prendre qu'un nombre limité de valeurs. ex : nombre de personnes dans un foyer)
- Variables Qualitatives
 - Variables qualitatives binaires (sexe : Masculin / Féminin)
 - Variables qualitatives nominales (Couleurs des yeux : marrons, bleus, verts, gris)
 - Variables qualitatives ordinales (Appréciation : Mauvais, Passable, Bien, Très bien, Excellent)

Toujours décrire les données avant de faire les analyses inférentielles (tests)

- Pour décrire les échantillons et vérifier leur représentativité mais aussi
- Pour le contrôle de qualité des données : individus aberrants, valeurs manquantes
- Pour choisir les tests adaptés aux distributions ("lois") des variables

Dans le cadre des méthodes statistiques inférentielles :

Une variable est définie par

- son type (quantitative, qualitative)
- son statut (++)

2 statuts possibles :

- Variables explicatives
 - \equiv variables indépendantes, variables exogènes, prédicteurs...
 - Variables dont on se sert pour expliquer le phénomène à l'étude.
 - Il s'agit de la cause présumée.
- Variable à expliquer
 - \equiv variable dépendante, variable endogène, critère de jugement...
 - Variable dont on veut expliquer la variation dans une recherche

Statut des variables - Exemple 1

Question d'étude : les enfants ayant eu des affections chroniques décèdent-ils plus en réanimation ?

Statut : Variables explicatives

- Admission et pendant séjour
 - surpoids : poids (kg)
 - indice de masse corporelle (poids/m²)
 - moyen de transport
 - présence d'affection chronique (oui/non)

Statut : Variable à expliquer

- Qualitative binaire (oui/non)
- Critère de jugement
- Sortie de réanimation
 - Décès (oui/non)

Statut des variables - Exemple 1

Question d'étude : les enfants ayant eu des **affections chroniques** décèdent-ils plus en réanimation ?

Statut : **Variables explicatives**

- Admission et pendant séjour
 - surpoids : poids (kg)
 - indice de masse corporelle (poids/m²)
 - moyen de transport
 - présence d'affection chronique (oui/non)

Statut : **Variable à expliquer**

- Qualitative binaire (oui/non)
- Critère de jugement
- Sortie de réanimation
 - Décès (oui/non)

Statut des variables - Exemple 1

Question d'étude : les enfants ayant eu des **affections chroniques** décèdent-ils plus en réanimation ?

Statut : **Variables explicatives**

- Admission et pendant séjour
 - surpoids : poids (Kg)
 - Index cardiaque : petit, moyen, gd
 - Immunodépression : oui/non
 - ...

Statut : **Variable à expliquer**

- Qualitative binaire (oui/non)
- Critère de jugement
- Sortie de réanimation
- Décès (oui/non)

Statut des variables - Exemple 1

Question d'étude : les enfants ayant eu des **affections chroniques** **décèdent-ils** plus en réanimation ?

Statut : **Variables explicatives**

- Admission et pendant séjour
 - surpoids : poids (Kg)
 - Index cardiaque : petit, moyen, gd
 - Immunodépression : oui/non
 - ...

Statut : **Variable à expliquer**

- Qualitative binaire (oui/non)
- Critère de jugement
- Sortie de réanimation
 - ⇒ **Décès (oui/non)**

Statut des variables - Exemple 2

Question d'étude : les enfants ayant eu des affections chroniques ont-ils une durée de séjour plus longue en réanimation ?

Statut : Variables explicatives

- Admission et pendant séjour
 - surpoids : poids (Kg)
 - le index cardiaque (petit)
 - moyen, etc.
 - les affections chroniques

Statut : Variable à expliquer

- Quantitative
- Critère de jugement
- Sortie de réanimation
- Durée de séjour

Statut des variables - Exemple 2

Question d'étude : les enfants ayant eu des **affections chroniques** ont-ils une durée de séjour plus longue en réanimation ?

Statut : **Variables explicatives**

- Admission et pendant séjour
 - surpoids : poids (Kg)
 - Index cardiaque : petit, moyen, gd
 - Immunodépression : oui/non
 - ...

Statut : **Variable à expliquer**

- Quantitative
- Critère de jugement
- Sortie de réanimation
- Durée de séjour

Statut des variables - Exemple 2

Question d'étude : les enfants ayant eu des **affections chroniques** ont-ils une **durée de séjour** plus longue en réanimation ?

Statut : **Variables explicatives**

- Admission et pendant séjour
 - surpoids : poids (Kg)
 - Index cardiaque : petit, moyen, gd
 - Immunodépression : oui/non
 - ...

Statut : **Variable à expliquer**

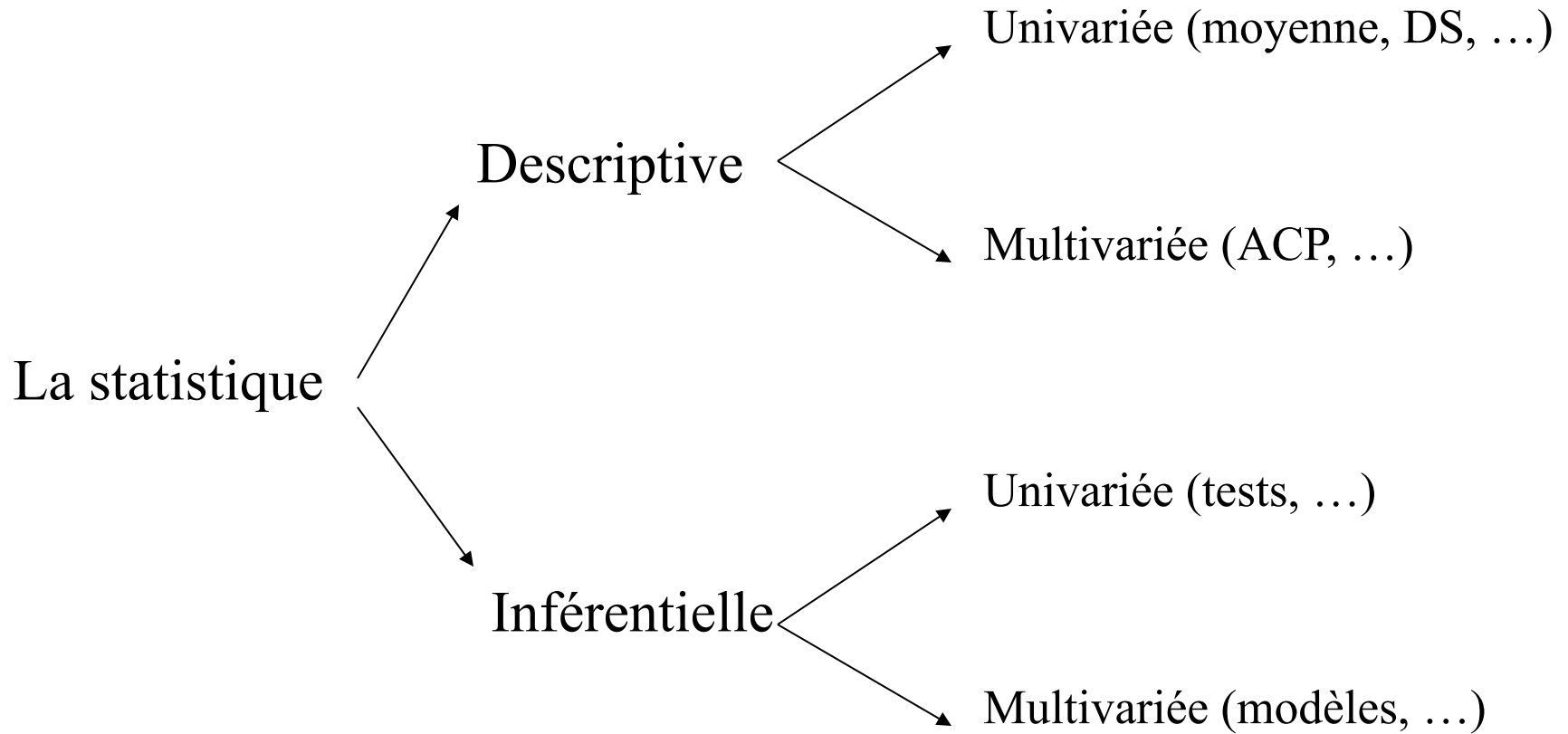
- Quantitative
- Critère de jugement
- Sortie de réanimation
⇒ **Durée de séjour (j)**

- 1) Les méthodes statistiques à employer dépendent toujours du type de variables à analyser
- 2) Pour les analyses descriptives :
 - On identifie le type de chaque variable
 - Toutes les variables ont-elles le même type, ou mélange ?
- 3) Pour les analyses inférentielles, on détermine en plus :
 - La variable à expliquer
 - Les variables explicatives

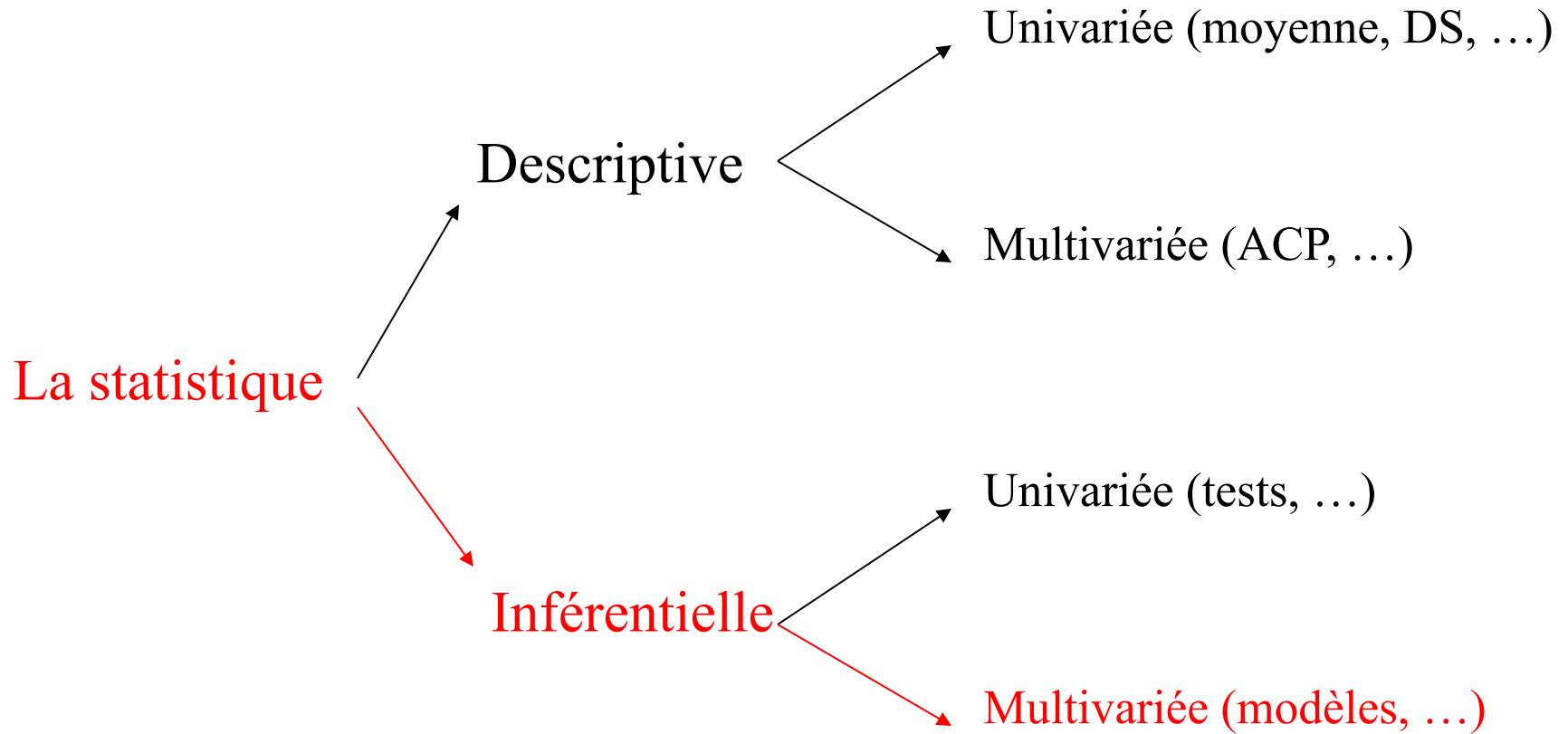
Plan de la présentation

- Objectif : Présenter une vue générale des modèles statistiques simples à utiliser en recherche clinique en médecine
 1. Enjeux
 2. Variables
 3. Méthodes
 4. Perspectives

Les méthodes statistiques



Les méthodes statistiques



Principales méthodes

Régression multiple

- Un critère numérique (variable à expliquer)
- Des variables explicatives numériques ou binaires

Analyse discriminante

- Un critère qualitatif (variable à expliquer)
- Des variables explicatives numériques ou binaires

Régression logistique

- Ajustement ou recherche de facteur de risque
- Un critère binaire
- Des variables explicatives numériques ou binaires

Mesures répétées

- les valeurs d'un critère (principal ou secondaire) sont relevées à différents temps

Analyses de survie (études pronostiques)

- Un événement qui peut se produire à un temps t
- Des variables explicatives numériques ou binaires

Les 3 grandes classes de modèle multivarié

- Régression linéaire multiple
 - La variable à expliquer est continue
 - Ex : albumine = $f(\text{poids, taille, apports nutritionnels, démence...})$
- Régression logistique
 - La variable à expliquer est catégorielle
 - Ex : démence (0/1) = $f(\text{niveau socio-éducatif, pression artérielle, médicaments...})$
 - Ex : décès dans les 3 mois (0/1) = $f(\text{confusion, marche, HTA, IC, IRC...})$
- Modèle de Survie (modèle de Cox)
 - La variable à expliquer est le temps écoulé avant événement (couple $\{T_i; D_i\}$)
 - Ex : temps écoulé avant décès = $f(\text{confusion, marche, HTA, IC, IRC...})$

- 1) Le modèle à utiliser dépend de la nature de la variable à expliquer
- 2) Il faut être très clair sur l'objectif de votre recherche
= quelle est la question posée ?

Plan de la présentation

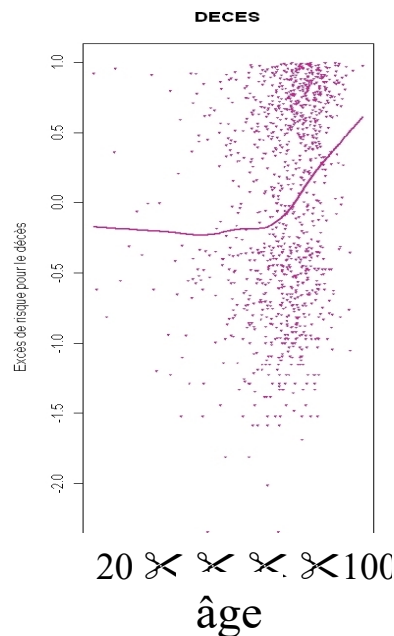
- Objectif : Présenter une vue générale des modèles statistiques simples à utiliser en recherche clinique en médecine
 1. Enjeux
 2. Variables
 3. Méthodes
 4. Perspectives

Aspects techniques

- Un modèle statistique multivarié décrit une association entre une variable à expliquer et des variables explicatives
- Choix du modèle : selon la nature de la variable à expliquer
- Variables explicatives à inclure dans le modèle initial:
 - Choix des variables initiales
 - Codage de variables continues
- Sélection des variables
 - Procédures automatiques
 - Stratégies de sélection
- Adéquation du modèle

Un modèle impose une « forme » à la réalité

- **Hypothèses du modèle utilisé** : il faut les connaître et les tester
- Exemple : hypothèse de log-linéarité
 - Modèle de Cox et modèle Logistique
 - Ex : Augmentation du risque de décès est identique pour 30 → 50 ans et 60 → 80 ans



Variable à expliquer : temps avant décès
Individus : patients en dialyse péritonéale
Population : Registre REIN
VA explicative : âge
Rejet de l'hypothèse de log-linéarité

Un modèle impose une « forme » à la réalité

- **Hypothèses du modèle utilisé** : il faut les connaître et les tester

Une analyse statistique multivariée de bonne qualité demande une analyse par un expert en statistique :

- Maîtrise du modèle
- Maîtrise de la programmation
- Test sur les variables
- Méthodes de sélection
- Adéquation du modèle
- Maîtrise de la rédaction (publication)
- Etc...

Bac +5 à +10 : c'est un métier!!!

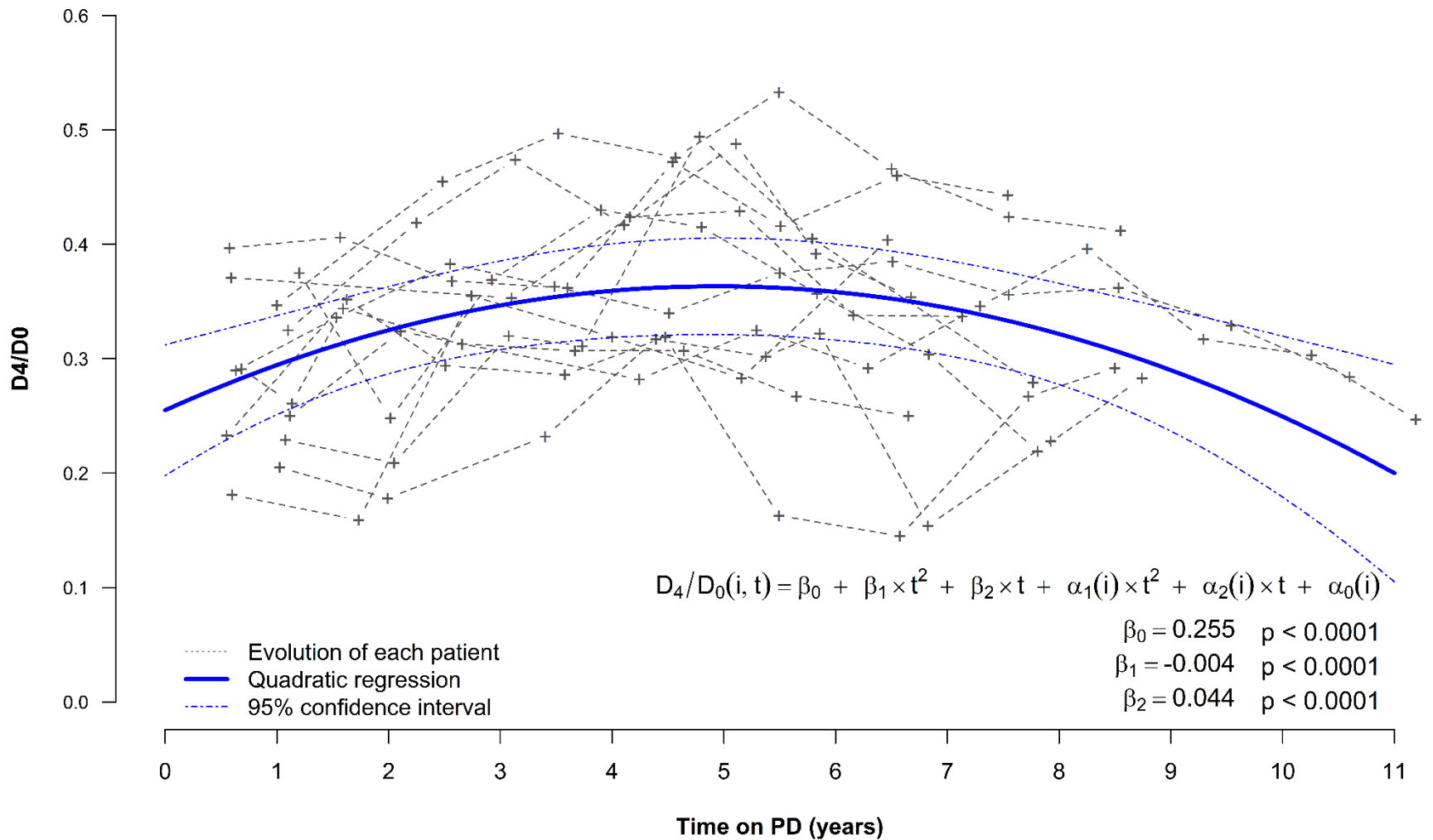
Modèle linéaire multiple : adéquation

Cours de M2 :
même pas d'expérience une
fois qu'on a appris ces items!

Partie 2

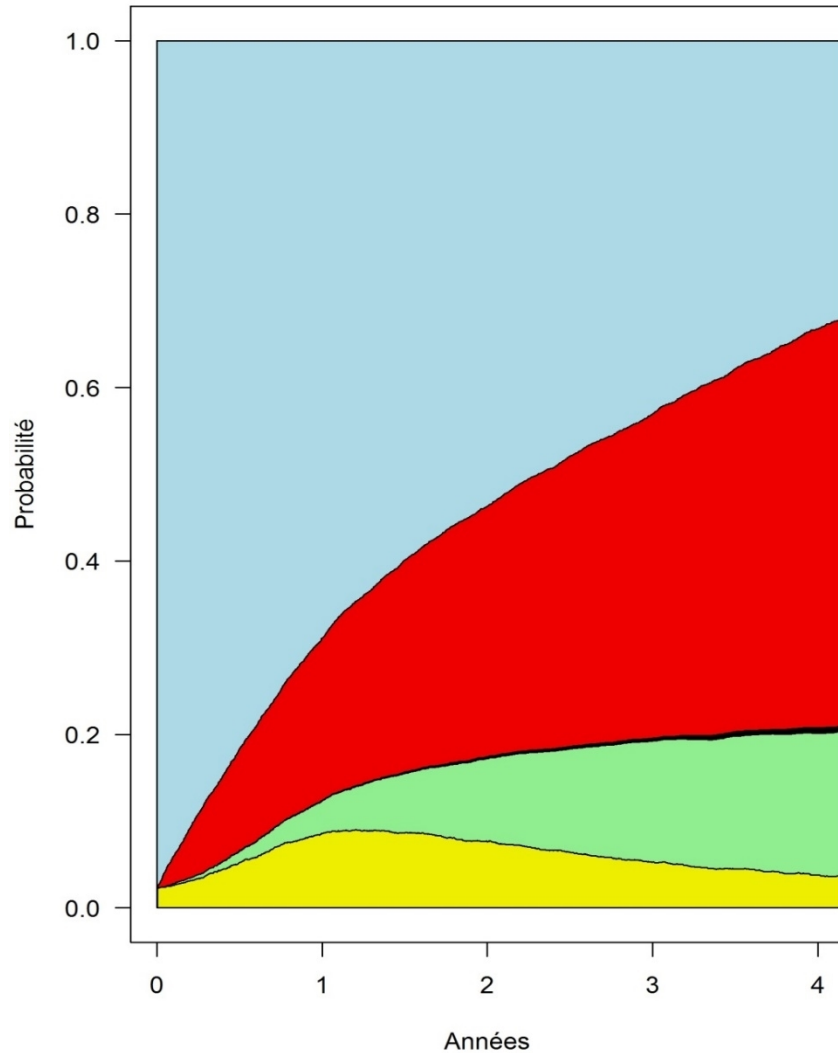
Outils diagnostiques	2
• éléments diagonaux de la matrice chapeau	8
• résidus	16
• mesure d'influence	27
• covratio	32
• dfbetas	35
• graphe de régression partielle	41
• que faire des observations influentes	50
Validité du modèle	52
• normalité	53
• homoscedasticité	62
• linéarité	67
Colinéarité	74
Sélection des variables	79

Le modèle peut être adapté à la questions / aux données

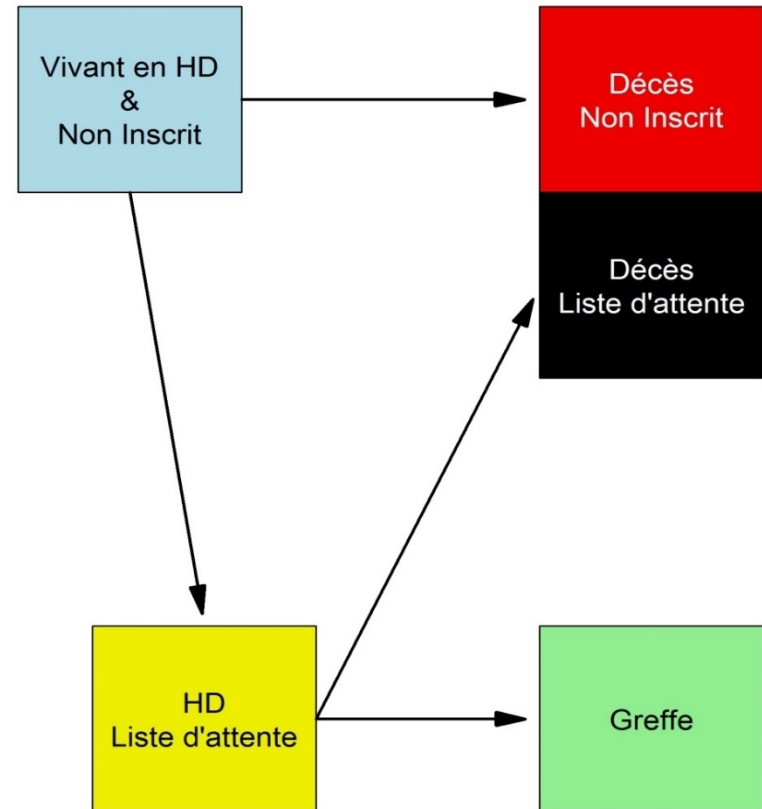


Beuscart et al. BMC Nephrol. 2017

La question du bon modèle peut être un sujet de recherche

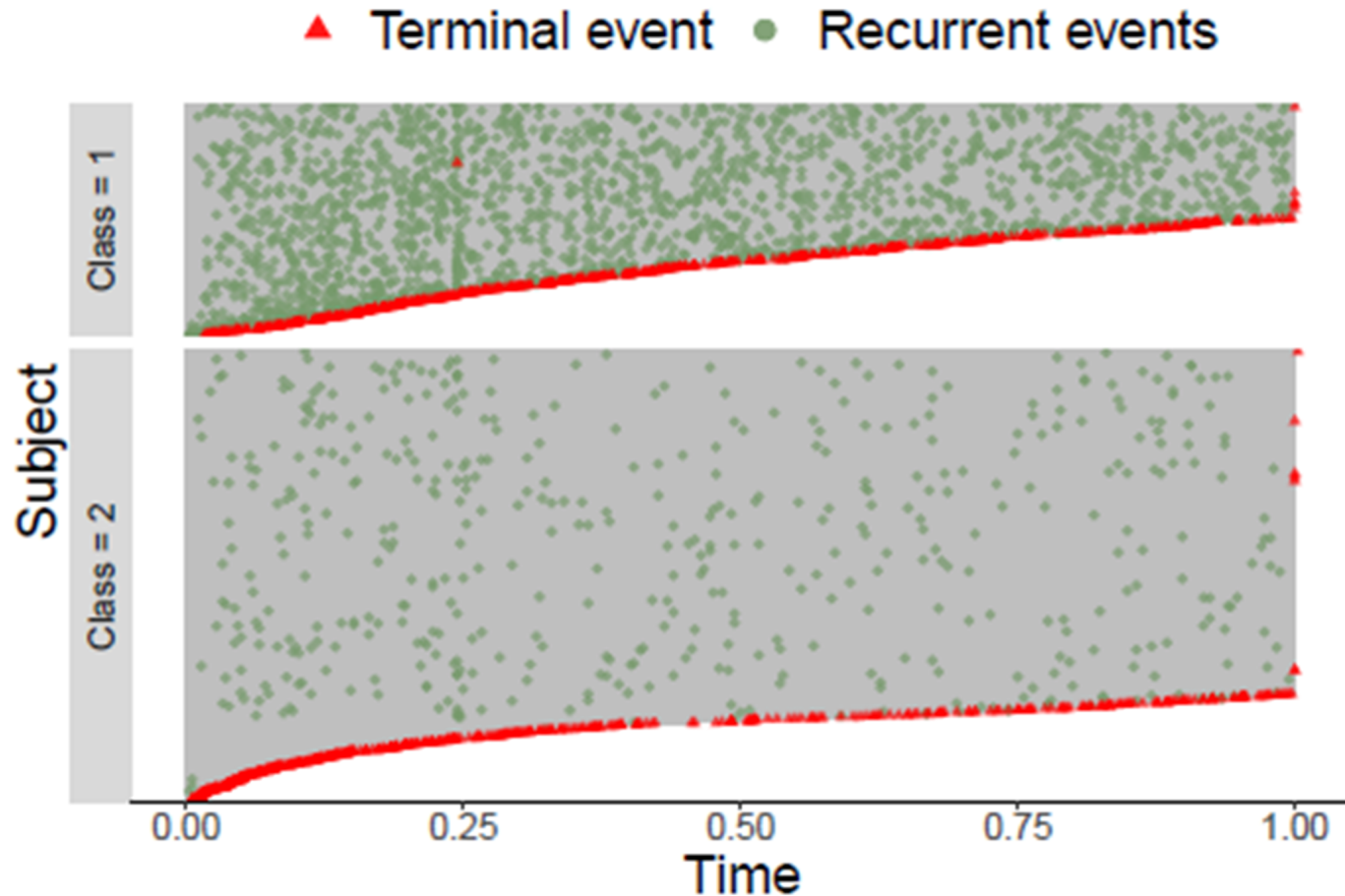


Légende :



Beuscart et al. J. Epidemiol. 2015

La question du bon modèle peut être un sujet de recherche



Latent class analysis, DAMAGE cohort, J. Carratero

Votre travail

AVANT de rencontrer les experts en statistique

- Bien faire la bibliographie sur le sujet
- Discuter de la question au vu de la bibliographie
- Essayer de formaliser la question (sur présentation PPT, rédaction minimale avec les VA à expliquer/explicatives, etc.)

Rencontre et échange avec les statisticiens

- 1^{ère} rencontre : projet, rationnel, question posée, VA, modèles retrouvés dans les articles de biblio, etc. => permet de déterminer le projet d'analyse statistique
- Travailler la rédaction du projet avec tous les acteurs (experts médicaux, méthodologistes, statisticiens, régulations locales, etc.)
- 2^{ème} rencontre (ou échanges mails) : validation analyse statistique, nombre de sujets nécessaires

Messages

- 1) Un modèle statistique impose une forme à la réalité décrite. Il faut donc vérifier si les hypothèses du modèle sont respectées.
- 2) Une analyse multivariée se fait de préférence avec un expert en statistique qui maîtrise le modèle utilisé
- 3) **Votre travail** : être très clair sur la population, les individus, la variable à expliquer, les variables explicatives... et la question posée!