

Science des Données en Santé

Introduction

Dr Thibaut FABACHER

GMRC
Service de santé publique
CHU Strasbourg
thibaut.fabacher@chru-strasbourg.fr

31 mars 2021

Plan de la section

- 1 Introduction
- 2 Historique
- 3 Science des données
- 4 Science des données en santé

Spécialiser en **Santé publique**

Master 2 : Biostatistique et données massives en santé

Thèse de science en cours en **TAL**

Fort intérêt pour la **science des données**

- ❶ 01/04 : Introduction
- ❷ 01/04 : Description des données de santé
- ❸ 08/04 : Visualisation des données de santé
- ❹ 15/04 : Techniques « explicatives » /modélisations statistiques
- ❺ 29/04 : Techniques prédictives / Apprentissage Machine
- ❻ 06/05 : Réseaux de neurones

- 1h de restitution
- 3h de cours tutoré le matin
- 3h de travaux personnels l'après midi

Outils nécessaires

- Logiciel R : <https://cloud.r-project.org/>
- Logiciel Rstudio : <https://rstudio.com/>
- Un compte google
- lien vers le github : <https://github.com/DrFabach/sciDoCours>

Plan de la section

- 1 Introduction
- 2 Historique
- 3 Science des données
- 4 Science des données en santé

Les débuts de la statistique

Présente depuis le début de l'humanité

- Recensement du bétail
- Suivi du cours des marchandises
- Recensement de la population (Rome Antique)

Premier livre



Al-Kindi (801-873)

Manuscrit pour Décrypter les messages cryptographiques

Description détaillée de la façon d'utiliser les statistiques et l'analyse de la fréquence de déchiffrer les messages chiffrés

Présent dans différents domaines scientifiques

- **Démographique**, John Graunt estime la population de Londres à partir des registres paroissiaux (1662)
- **Astronomie**, méthode des moindres carrés pour calculer la trajectoire de Cérès (1801)
- **Physique** (Thermodynamique)

La statistique permet de comparer des données expérimentales à un modèle mathématique censé décrire ces données.

Fondateur de la statistique moderne :

- **Laplace**, *Théorie analytique des probabilités*, 1812, Théorème central limite
- **Bayes**, théorème

- **Science à part entière**, 1919, première chaire de statistique à l'université de Londres
- Début des applications industrielles
- **Avènement de l'ordinateur en 1940**

La statistique, une discipline née dans un monde sans informatique et pauvre en données.

11 12 13 14 15
 16 17 18 19 20

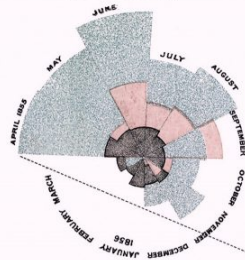
01-N.D.0447

Les statistiques descriptives et exploratoires

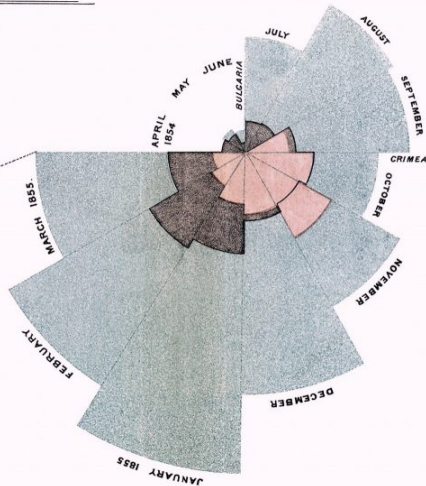
- Résumés, représentations graphiques
- Décrire
- Mettre en évidence des relations

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Les statistiques inférentielles et confirmatoires

- Expliquer
- Prédire

- Fréquentiste : repose sur la loi des observations
 $\text{prob}(\text{données}|\text{hypothèse})$
- Bayésianisme : combine l'information apportée par les données avec les connaissances a priori
 $\text{prob}(\text{hypothèse}|\text{données})$

<https://www.youtube.com/watch?v=x-2uVNze56s>

- **1956** : Naissance de l'intelligence artificielle
- 1958 : Perceptron

1965, H. Simon :

« des machines seront capables, d'ici vingt ans, de faire tout travail que l'homme peut faire »

- Années 70 : IA Winter
- Années 80 : Système Expert
- Années 2010 : Retour en force du deeplearning : IBM **Watson** (Jeopardy), **Alpha GO**

Plan de la section

1 Introduction

2 Historique

3 Science des données

4 Science des données en santé

Statistique, IA, Science des données ?

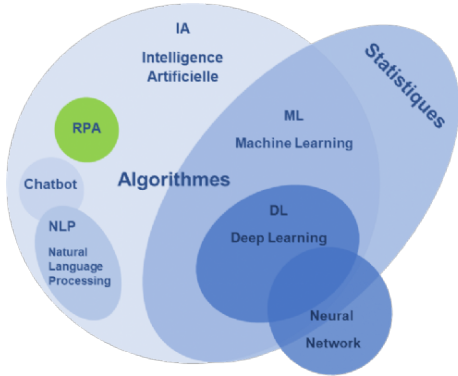
Statistique

"Ensemble des méthodes qui ont pour objet la collecte, le traitement et l'interprétation de ces données" ¹

IA

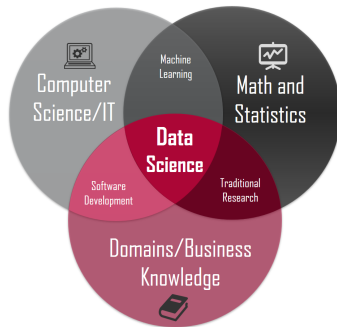
"Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine" ¹

Statistique, IA, Science des données ?



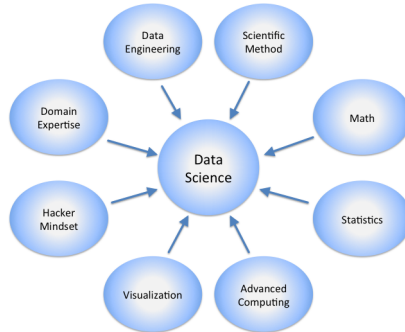
1. <https://www.daf-mag.fr/Thematique/fonction-finance-1242/Breves/Les-Deeptech-secours-Controle-Gestion-quelles-applications-Machine-Learning-RPA-342807.htm>

Statistique, IA, Science des données ?



1. <https://towardsdatascience.com/why-data-science-succeeds-or-fails-c24edd2d2f9>

Statistique, IA, Science des données ?



1. <https://fr.wikipedia.org/>

Plan de la section

- 1 Introduction
- 2 Historique
- 3 Science des données
- 4 Science des données en santé**

- Recherche clinique : **Statistique**
- Génomique : **Big Data**
- Imagerie : **Réseaux de neurones**
- Dossier médical : **TAL**
- Epidémiologie, Cartographie, Médecine prédictive, Médecine personnalisée, Médicaments ...