

Tout Bachelier

Éléments de Statistique

Enseignement supérieur économique de type court

Code FWB : 013203U32D2

Code ISFCE : 4XSTA

Le Professeur, c'est qui ?

Alain Wafflard alain.wafflard@gmail.com

Ingénieur civil

25 ans dans le privé
Processus d'entreprise
Automatisation, informatisation

Enseignant

ISFCE Etterbeek
mais auparavant : IFOSUP Wavre, IFC Ixelles, IEPSCF Evere, IPAM La Louvière, EPFC Bruxelles



Activité complémentaire

burotix.be
formation en tableur
conception de base de données
développement de templates

Statistique Descriptive
Chapitre 00 : Introduction

A propos du cours

Finalité du cours

- s'approprier des connaissances générales en statistique en vue de traiter des données économiques, financières, commerciales, ….
- développer son sens critique dans l'interprétation de données ou documents statistiques
- utiliser des ressources logicielles existantes permettant le traitement statistique de données

Programme du cours

- Concepts de base de la statistique descriptive (chap. 00)
- Éléments de statistique univariée (chap. 01)
- Éléments de statistique bivariée (chap. 02)

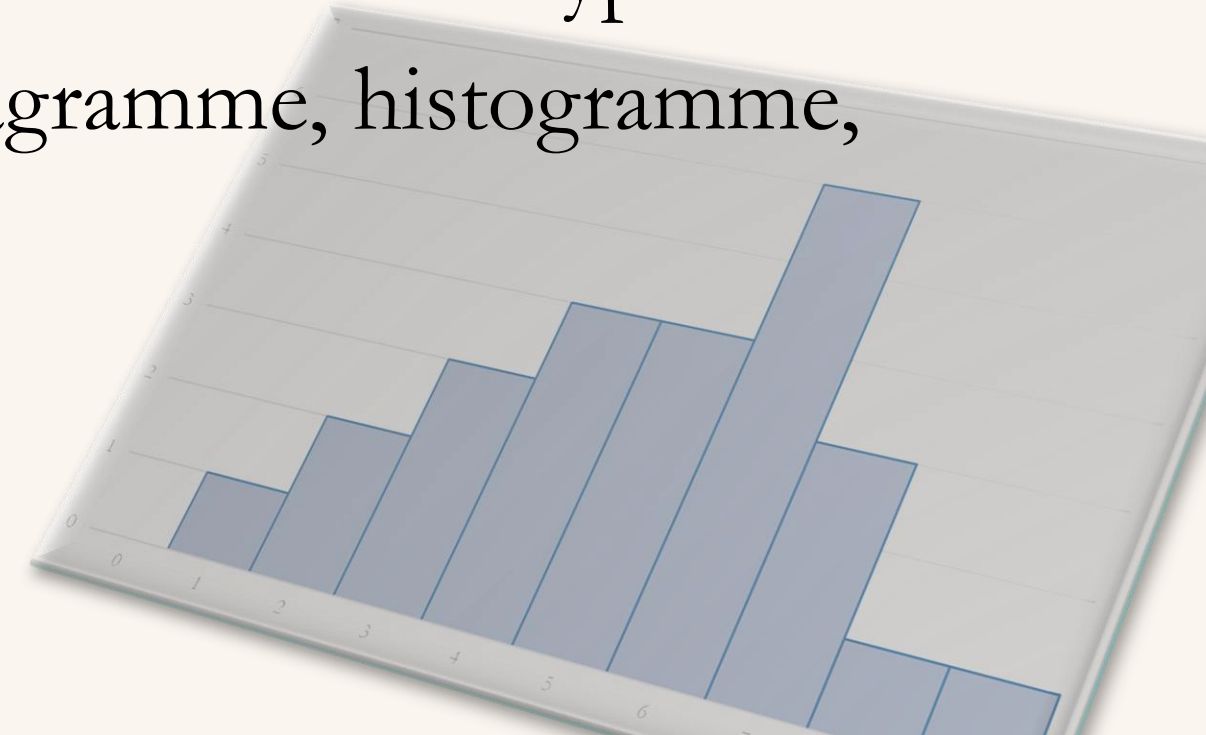
Chapitre 00 : Concepts de base de la statistique descriptive

- (ce chapitre-ci ☺)
- Historique
- Définitions
- Paradigmes



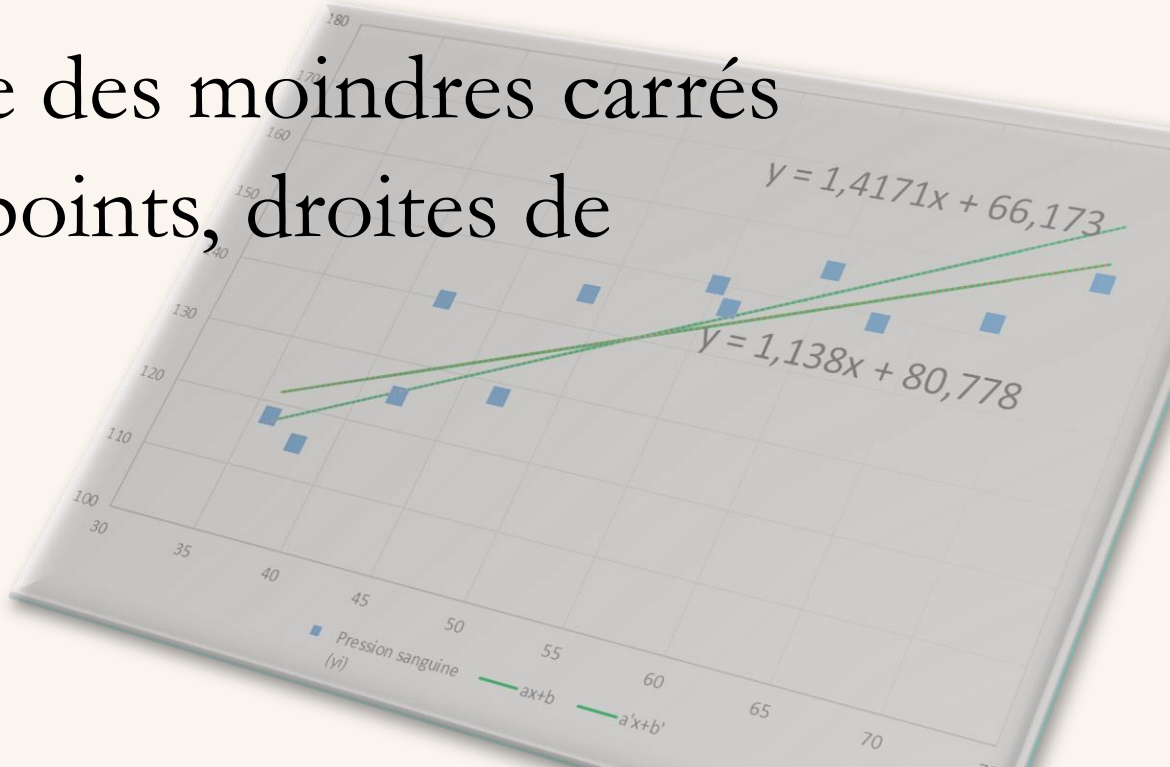
Chapitre 01 : Statistique Univariée

- Caractère quantitatif discret ou continu
- Fréquence absolue ou relative, simple ou cumulée
- Médiane, Mode, Moyenne, variance et écart-type
- Représentation graphique (diagramme, histogramme, etc.)
- Interprétation
- Traitement automatisé



Chapitre 02 : Statistique Bivariée

- Nuage de points et point moyen "g"
- Indépendance et dépendance fonctionnelle
- Corrélation
- Ajustement linéaire et méthode des moindres carrés
- Ajustement linéaire, nuage de points, droites de régression
- Interprétation
- Traitement automatisé



Définitions

un peu d'histoire

définition générale

paradigmes

Population et individus

Variables et modalités

Types de variables

Échantillon

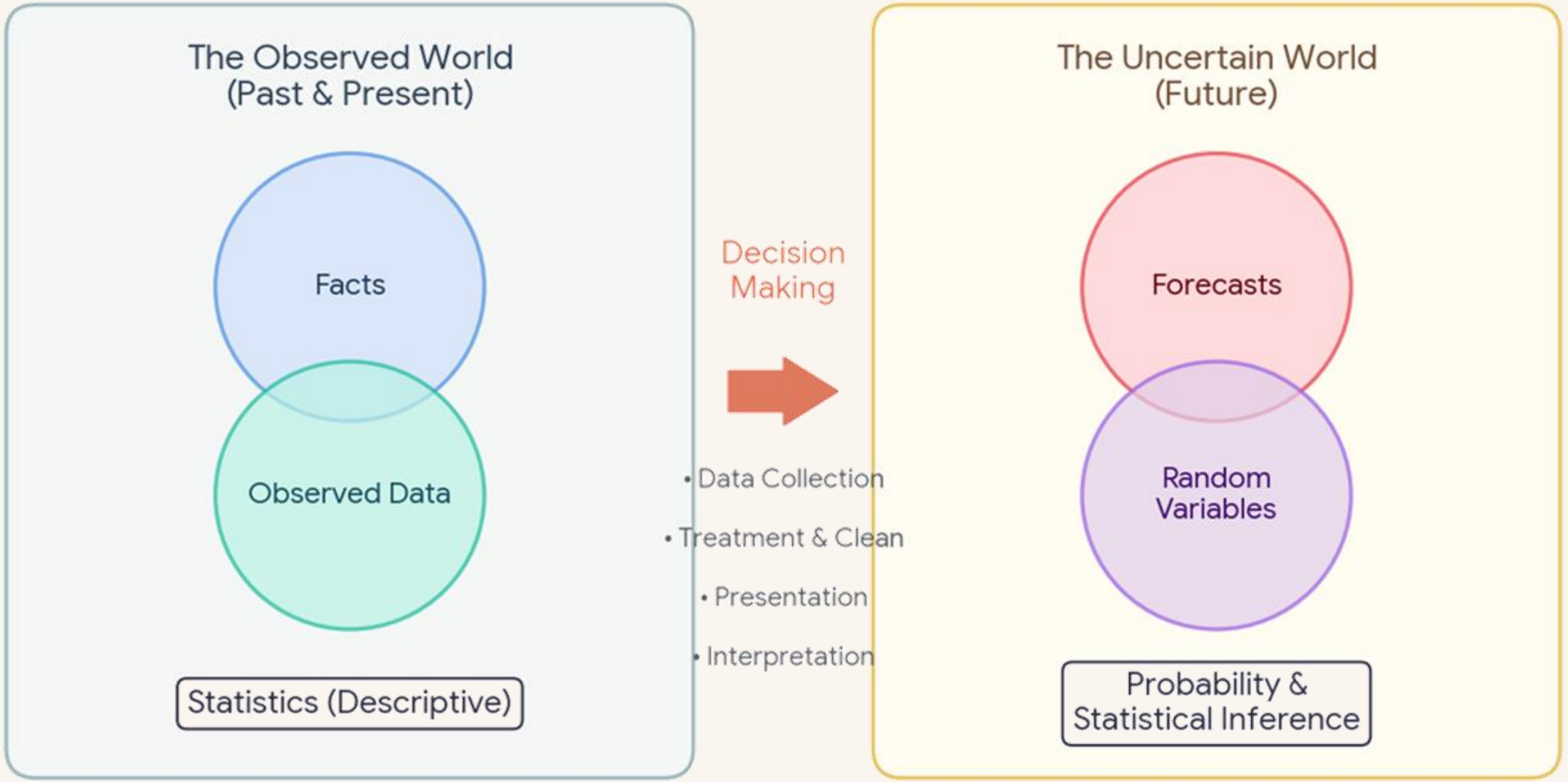
Un peu d'histoire

- recensement du bétail et de son prix
- origine moderne : statista
  - "ensemble des connaissances que doit posséder un homme d'État"
    - cf ouvrage de Gottfried Achenwall : *"Abriss der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republiken zum Gebrauch in seinen Akademischen Vorlefungen"*, 1749
- rôle prévisionnel
  - assurances

Statistique : Définition

- Ensemble des méthodes qui ont pour objet, sur base de données d'observation sur un groupe d'individus
  - la collecte des données à observer
  - leur traitement et leur analyse
  - leur présentation, visant à les rendre compréhensibles
  - l'interprétation des résultats en vue d'une prise de décision

Data Science Framework: Descriptive vs. Inferential Horizons



Unified Analytical Horizon

Statistique : Définition

- Exemple
  - Un professeur souhaite étudier les habitudes de lecture de ses étudiants.
  - Collecte des Données
    - Questionnaire distribué à 200 étudiants.
  - Données recueillies
    - Nombre de livres lus par mois.
    - Genre préféré (roman, …)
  - Traitement et Analyse
    - Outil : Tableur
    - Calculs : Moyenne du nombre de livres lus par mois.
- Présentation des Résultats
  - Genre Pourcentage
    Roman 45%
    Scientifique 20%
    BD 30%
- Interprétation
  - La majorité des étudiants lisent des romans.
  - 30% préfèrent les BD.
  - Recommandation : Enrichir la bibliothèque avec plus de BD.

Statistique **Descriptive** : Définition

- Ensemble des techniques permettant de
  - Résumer et présenter les données
  - Faciliter leur compréhension
  - Représenter les données sous forme de tableaux, de graphiques et d'indicateurs numériques.

… quand les données disponibles sont nombreuses et complètes.

Statistique Inférentielle : Définition

Ensemble des techniques permettant de

- induire (prédire) les caractéristiques d'une population à partir de celles d'un échantillon,
  - en estimant la (in)certitude de la prédiction.
- tester si les données d'un échantillon corroborent bien une hypothèse de départ concernant une population.

Dit autrement,

- L'échantillon est traité d'une manière exacte (statistique descriptive).
- Puis, les résultats sont élargis à l'échelle de la population, en estimant l'incertitude (statistique inférentielle).

Paradigmes statistiques : Trois façons d'interpréter la probabilité

La disponibilité et le coût des données orientent le choix de l'approche.

01 Fréquentisme
OBJECTIVISME

IDÉE-CLÉ

La répétition révèle la probabilité par la fréquence observée.

EXEMPLE

10 000 lancers

La fréquence de « pile » s'approche de 0,5.

→ **P(pile) = 0,5**

02 Subjectivisme
CONFIANCE

IDÉE-CLÉ

La probabilité traduit un degré de confiance révisable.

EXEMPLE

Une élection

« Quelle est la probabilité que ce candidat gagne l'élection ? »
Pas de répétition possible : le chiffre reflète notre confiance, à réviser avec l'expérience.

03 Bayésisme
APPRENTISSAGE

IDÉE-CLÉ

Chaque nouvelle donnée met à jour le poids des hypothèses.

EXEMPLE

Plusieurs diagnostics

En médecine, chaque résultat d'examen renforce ou affaiblit chacune des hypothèses.

Variables qualitatives et quantitatives

- Variable qualitative (ou "nominales"):
  - Elle exprime une qualité.
  - Elle se traduit par des noms:
    - couleurs : rouge, vert, bleu, …
    - groupe sanguin : A+, A-, AB+, …
  - On ne peut les traiter mathématiquement, on ne peut la traduire en quantité.

$$x = m_1, m_2, \ldots, m_q$$

Variables qualitatives et quantitatives

- Variable quantitative:
  - Elle exprime une quantité.
  - Elles sont mesurables.
    - Taille
    - Poids
    - Notes d'un examen
  - Elles sont traitables numériquement.

$$x_i = 1, 2, 3 \ldots n$$

Variables qualitatives et quantitatives

- Deux types de variable quantitative :
  - la variable quantitative discrète
    - nombre fini de valeurs possibles
    - exemple
      - nombre de personnes vivant dans mon domicile
  - la variable quantitative continue
    - toutes les valeurs sont possibles dans un intervalle [a,b].
    - exemple
      - Poids, taille, revenu, …

Population et individus

- Individu ou unité statistique
  - Une unité distincte chez laquelle on peut observer une ou plusieurs caractéristiques données
- Population
  - Ensemble des individus (ou unités statistiques) pour lequel on a au-moins une propriété en commun.
- Échantillon
  - C'est un sous-ensemble de la population.
  - L'échantillon doit être représentatif de la population.

- Exemple
  - On veut étudier le niveau d'études de la population bruxelloise adulte
  - unité
    = la personne adulte habitant Bruxelles
    = 1 personne
  - population
    = l'ensemble des Bruxellois adultes
    = 1 200 000 personnes
  - échantillon
    = sous-ensemble des Bruxellois adultes représentatifs de la population
    = 2000 personnes

Quelques références

livres, articles

site web, chaînes YouTube

logiciels

Livres

- Daudin Jean-Jacques, Robin Stéphane, Vuillet Colette,
  Statistique inférentielle - Idées, démarches, exemples
  Presses Universitaires de Rennes, 1999
  EAN : 9782868474049
- Nicolas Gauvrit
  Statistiques : Méfiez-vous !
  Éditions ellipses, 2007
  EAN : 9782729884468

Sites Web - chaînes YouTube

- Wikipedia
  - https://fr.wikipedia.org/
- JMP Software
  - https://www.jmp.com/
- Khan Academy
  - https://fr.khanacademy.org/
- Fabinou
  - https://fabinou.fr/

Autres Cours & Syllabus

- Jonathan Lenoir, Université de Picardie
  - introduction à la statistique inférentielle

Logiciels de statistiques

- Qualtrics Stats iQ.
  - https://www.qualtrics.com
- XLSTAT Cloud
  - https://lumivero.com/
- Tanagra, Add-In for Excel, logiciel libre
  - Ricco Rakotomalala
  - https://tanagra.software.informer.com/
- R, logiciel libre
  - R Core Team
  - https://www.r-project.org/