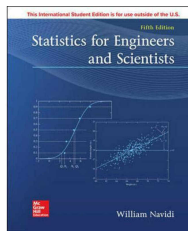
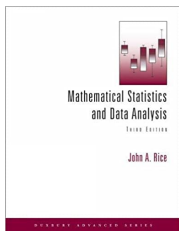


# Introduction aux statistiques pour l'Ingénieur

Stéphane Canu

[asi.insa-rouen.fr/enseignants/~scanu](http://asi.insa-rouen.fr/enseignants/~scanu)

[scanu@insa-rouen.fr](mailto:scanu@insa-rouen.fr)



ITI 3, INSA Rouen Normandie, Janvier 2024

# Statistiques pour l'ingénieur

Tout est sur moodle (n'oubliez pas de vous inscrire)

▼ Tableau de bord

🏠 Accueil du site

➤ Pages du site

▼ Mes cours

- Admin-STP-enseignants
- Admin-TI
- Admin-TI-enseignants
- ITI-apprentissage-en-contexte
- TDGROASI
- ASI-DM
- ASI-FaUILConnex2
- ASI\_C1
- PAD
- Stages-ITI
- ▼ ASI-Stats
- Participants
- 📄 Notes

➤ Bienvenue au cours de statistiques pour l'ingénieur !

- Section 1
- Section 2
- Section 3
- Section 4
- Section 5
- Section 6
- Section 7
- Section 8
- Section 9
- Section 10
- Section 11
- Section 12
- Admin-MS-ESD
- lasso\_stat2
- Machine\_Learning\_avancé
- ML1 et 2
- ML3-Deep-Learning
- RAN
- Système\_recommandation
- Comité VSS

Administration

▼ Administration du cours

- 🔗 Paramètres
- Utilisateurs
- ▼ Filtres

## Bienvenu au cours de statistiques pour l'ingénieur !

Le but de cette U.V. est de se familiariser avec les raisonnements en présence d'alias, tout en présentant les bases mathématiques et les méthodes statistiques que l'ingénieur pourra être amené à rencontrer lors de sa vie professionnelle.

### Prérequis :

Les prérequis sont les notions de **probabilités** et de **mathématiques** enseignées au niveau Bac +2 (fonctions de plusieurs variables, dérivation, intégration).

### Évaluation :

- examen médian 30 %
- examen final 40 %
- projet 20 %
- TP 10 %

### Quelques informations utiles :

-  Forum de l'UV Statistiques
-  Sujets TD Stat 2023
-  Correction TD Stat 2020
- [Cliquez pour les évaluer](#)
-  le livre du cours : Mathematical Statistics and Data Analysis
-  Ressources pédagogiques en ligne ...
-  Rappels et exercices variables discrètes
-  Rappels et exercices variables à densité
-  Annonces

## Section 1

### EVALUATION

-  TP 1
-  TP 2
-  TP3
-  TP4
-  TP 5

## Section 2

### Fonctionnement du projet

Cette année, le projet est constitué d'une liste de points précis non abordés en cours.

Chaque binôme devra choisir l'un de ces points et préparer une présentation de 10 minutes et (éventuellement) une démo en python illustrant le principe et les résultats.

Chaque binôme sera évalué en fonction des réponses aux questions de l'enseignant, du contenu et de la forme de la présentation, ainsi que de la démo/application réalisée.

# Lecture road map

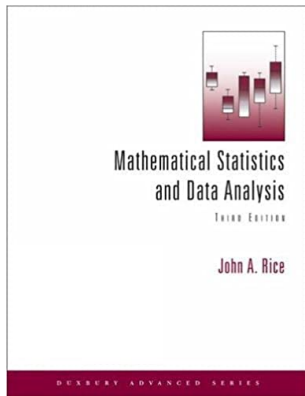
## 1 Statistiques pour l'ingénieur

- Mots pour mots
- Statistiques pour décrire
- Statistiques pour Inférer

## 2 Une brève histoire des statistiques

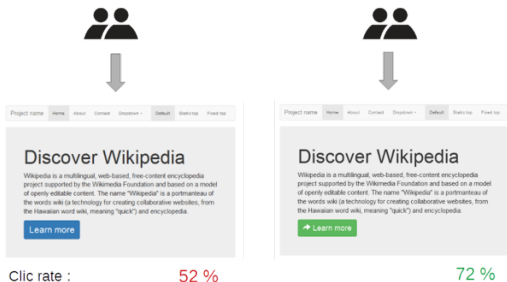
- Statistiques et état
- Statistiques et astronomie
- Statistiques et sciences

## 3 Enjeux actuels des statistiques



# Statistiques pour l'ingénieur

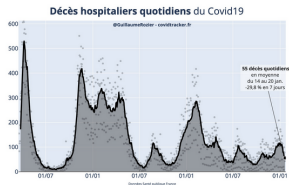
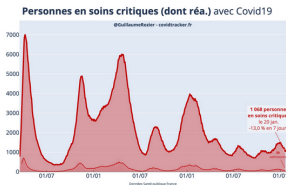
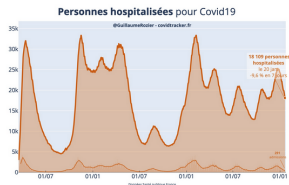
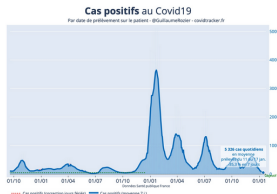
Pour l'ingénieur : aspect utile pour aider à prendre des décisions  
exemple : Le test A/B (ou A/B testing)



# Statistiques pour l'ingénieur

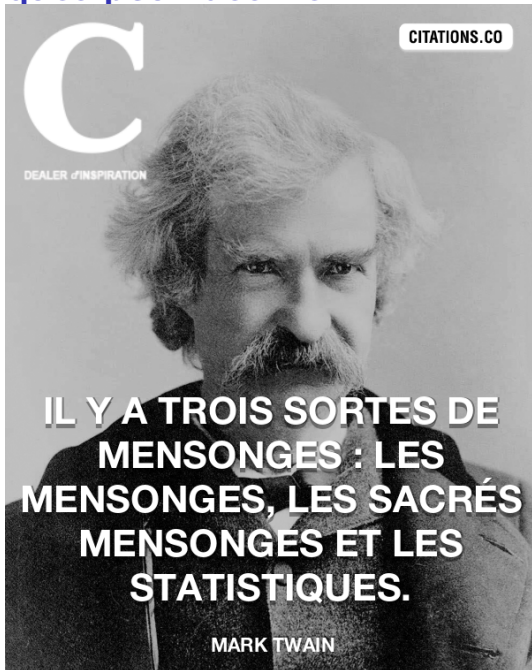
Statistiques : les statistiques sont partout

COVID : ou en est-on de la pandémie et ou va t'on ?



Décirie puis Décider

# Les statistiques pour décrire



# Statistiques pour décider

Applications :

- Evaluation des risques (assurance, dimensionnement de digues...)
- Statistique Biomédicale - Essais cliniques
- Détection d'anomalies
- Validation de Procédés - Fiabilité
- Enquêtes Sondage marketing
- Prévion météo (Séries Temporelles)
- ...

Modèle statistique :

- choix du modèle
- identification des paramètres
- prédiction/décision

$X$  suit une loi de Bernouilli

$$\hat{p} = \dots 51.5\%$$

je pense que ce candidat va être élu

# Lecture road map

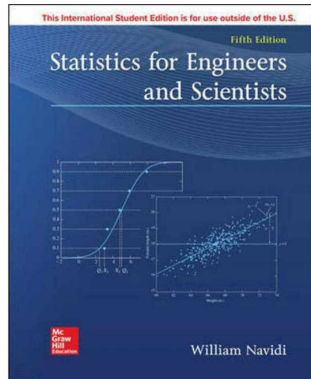
## 1 Statistiques pour l'ingénieur

- Mots pour mots
- Statistiques pour décrire
- Statistiques pour Inférer

## 2 Une brève histoire des statistiques

- Statistiques et état
- Statistiques et astronomie
- Statistiques et sciences

## 3 Enjeux actuels des statistiques





# Une brève histoire des statistiques

- Etymologie : latin classique status (état) → statistique en 1749.
- Statistik en allemand, *political arithmetic* en Angleterre jusqu'en 1798
- Organisation des bureau de statistiques : population (nombre de naissances/nombre de décès, richesse du pays...)

N<sup>o</sup>. 4. *Partie Troisième*. Pag. 309

## TABLE GENERALE.

Pour la Population de tout un Royaume, ou autre Etat quelconque, tirée des Tables Provinciales faites sur le modèle de la Table N<sup>o</sup>. 1. ci-dessus, par exemple, pour l'ANNEE 1759.

### L A F R A N C E.

Mo. de l'Année	Normandie		Picardie		Dauphiné		Langue doc		Flandres		Total pour tout le Royaume (en 1759) - Grande Bretagne, etc.
	Mo.	Mo.	Mo.	Mo.	Mo.	Mo.	Mo.	Mo.	Mo.		
Janv.	12	12	12	12	12	12	12	12	12	12	12
Févr.	12	12	12	12	12	12	12	12	12	12	12
Mars	12	12	12	12	12	12	12	12	12	12	12
Avril	12	12	12	12	12	12	12	12	12	12	12
Mai	12	12	12	12	12	12	12	12	12	12	12
Juin	12	12	12	12	12	12	12	12	12	12	12
Juillet	12	12	12	12	12	12	12	12	12	12	12
Août	12	12	12	12	12	12	12	12	12	12	12
Sept.	12	12	12	12	12	12	12	12	12	12	12
Oct.	12	12	12	12	12	12	12	12	12	12	12
Nov.	12	12	12	12	12	12	12	12	12	12	12
Déc.	12	12	12	12	12	12	12	12	12	12	12

### RECAPITULATION.

Mo. de l'Année	Augmentés	Diminuis	Sur ou sous le plus de six de six	Sur ou sous le plus ou moins de six de six	Sur ou sous le plus ou moins de six de six	Sur ou sous le plus ou moins de six de six
Janv.	12	12	12	12	12	12
Févr.	12	12	12	12	12	12
Mars	12	12	12	12	12	12
Avril	12	12	12	12	12	12
Mai	12	12	12	12	12	12
Juin	12	12	12	12	12	12
Juillet	12	12	12	12	12	12
Août	12	12	12	12	12	12
Sept.	12	12	12	12	12	12
Oct.	12	12	12	12	12	12
Nov.	12	12	12	12	12	12
Déc.	12	12	12	12	12	12

Tous les détails que l'on peut désirer des cette Table se trouvent déjà dans la Table N<sup>o</sup>. 1. ci-dessus, ou les Mémoires, par où l'on voit ce qui est en blanc.

1606.

### A TABLE of the CHRISTENINGS and MORTALITY

For the Year 1605 and 1606.

Weeks	Days of the Month	Chr.	Mo.	Pa.	Weeks	Days of the Month	Chr.	Mo.	Pa.
1	Jan 1	100	100	100	1	July 1	100	100	100
2	Jan 2	100	100	100	2	July 2	100	100	100
3	Jan 3	100	100	100	3	July 3	100	100	100
4	Jan 4	100	100	100	4	July 4	100	100	100
5	Jan 5	100	100	100	5	July 5	100	100	100
6	Jan 6	100	100	100	6	July 6	100	100	100
7	Jan 7	100	100	100	7	July 7	100	100	100
8	Jan 8	100	100	100	8	July 8	100	100	100
9	Jan 9	100	100	100	9	July 9	100	100	100
10	Jan 10	100	100	100	10	July 10	100	100	100
11	Jan 11	100	100	100	11	July 11	100	100	100
12	Jan 12	100	100	100	12	July 12	100	100	100
13	Jan 13	100	100	100	13	July 13	100	100	100
14	Jan 14	100	100	100	14	July 14	100	100	100
15	Jan 15	100	100	100	15	July 15	100	100	100
16	Jan 16	100	100	100	16	July 16	100	100	100
17	Jan 17	100	100	100	17	July 17	100	100	100
18	Jan 18	100	100	100	18	July 18	100	100	100
19	Jan 19	100	100	100	19	July 19	100	100	100
20	Jan 20	100	100	100	20	July 20	100	100	100
21	Jan 21	100	100	100	21	July 21	100	100	100
22	Jan 22	100	100	100	22	July 22	100	100	100
23	Jan 23	100	100	100	23	July 23	100	100	100
24	Jan 24	100	100	100	24	July 24	100	100	100
25	Jan 25	100	100	100	25	July 25	100	100	100
26	Jan 26	100	100	100	26	July 26	100	100	100
27	Jan 27	100	100	100	27	July 27	100	100	100
28	Jan 28	100	100	100	28	July 28	100	100	100
29	Jan 29	100	100	100	29	July 29	100	100	100
30	Jan 30	100	100	100	30	July 30	100	100	100
31	Jan 31	100	100	100	31	July 31	100	100	100

\* See's London's Remembrance.

A TABLE

L'art de gouverner : bien connaître pour bien administrer

# Une brève histoire des statistiques

- Recensement

- ▶ L'empereur chinois Yao, organisant le recensement des productions agricoles en 2238 avant J.-C.
- ▶ Cadastre chez les Egyptiens 1700 av J.-C.

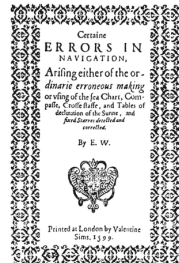
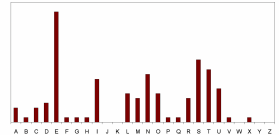
- Le premier écrit sur les statistiques :

« Manuscrit pour décrypter les messages cryptographiques », du mathématicien Al-Kindi (801-873) : la **fréquence**

- L'idée de **médiane** apparait dans un livre d'Edward Wright sur la navigation (Erreurs Certaines en navigation, 1599) dans une section concernant la détermination de sa localisation à la boussole.



Document 2 : diagramme en bâtons des fréquences d'apparition de chaque lettre de l'alphabet



# Utilité et statistique

Décision rationnelle en présence d'aléa :

- le pari de pascal
- gain (croire) > gain (ne pas croire)

PARI DE PASCAL	Dieu existe	Dieu n'existe pas
Vous pariez sur l'existence de Dieu	Paradis	Néant
Vous pariez sur la non existence de Dieu	Enfer ou Néant	Néant

Lady Montagu et l'inoculation préventive

- la variolisation : pré-vaccin de la variole
- risque = 0,5 à 2 % de mortalité
- premier débat sur les probabilités
- Sur quoi raisonner ?

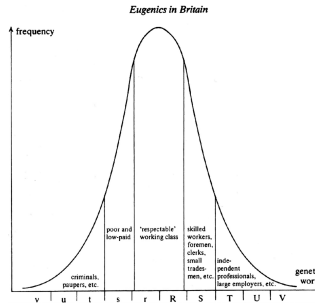
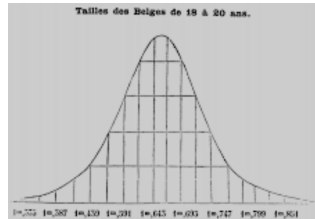
Résultat : iatrogénie



L'utilité est une question de point de vue

# Dérives : de l'homme moyen à l'eugénisme

- Quételet (1844-1848) : l'homme moyen (physique, social et moral)
  - ▶ la **normalité**
  - ▶ pour lui la moyenne c'est la perfection (???)
- l'eugénisme : déterminer les conditions d'amélioration de soit disant races → donner des bases scientifique au racisme
  - ▶ mesurer
  - ▶ définir ce que « mieux » veut dire
- des statisticiens importants étaient également des défenseurs passionnés de l'eugénisme.
- ce n'est pas neutre : En termes de statistiques de Pearson et Fisher, il existe un fort biais contre l'explication causale.
- certaines méthodes statistique et la pensée eugéniste sont, en fait, profondément liées,



# Rationalité des décisions

Doit-on jouer au loto (ou s'assurer) ?

- du point de vue de la française des jeux : c'est la moyenne qui permet de fixer les prix
- de notre point de vue, ce sont les écarts à la moyenne qui nous font acheter un billet

Fumer provoque-t-il le cancer ?

C'est toujours possible, mais est-ce probable ou même plausible ?

prouvé > possible > probable > plausible

L'astronome et le juge : Probabilité et fréquence objective (liée à un phénomène physique) ou degré de certitude subjective ou de croyance (l'intime conviction juridique)

# Choisir ou combiner ?

Table 1.1. Mayer's twenty-seven equations of condition, derived from observations of the crater Manilius from 11 April 1748 through 4 March 1749.

Eq. no.	Equation	Group
1	$\beta - 13^{\circ}10' = +0.8836\alpha - 0.4682\alpha \sin \theta$	I
2	$\beta - 13^{\circ}8' = +0.9996\alpha - 0.0282\alpha \sin \theta$	I
3	$\beta - 13^{\circ}12' = +0.9899\alpha + 0.1421\alpha \sin \theta$	I
4	$\beta - 14^{\circ}15' = +0.2221\alpha + 0.9750\alpha \sin \theta$	III
5	$\beta - 14^{\circ}42' = +0.0006\alpha + 1.0000\alpha \sin \theta$	III
6	$\beta - 13^{\circ}1' = +0.9308\alpha - 0.3654\alpha \sin \theta$	I
7	$\beta - 14^{\circ}31' = +0.0602\alpha + 0.9982\alpha \sin \theta$	III
8	$\beta - 14^{\circ}57' = -0.1570\alpha + 0.9876\alpha \sin \theta$	II
9	$\beta - 13^{\circ}5' = +0.9097\alpha - 0.4152\alpha \sin \theta$	I
10	$\beta - 13^{\circ}2' = +1.0000\alpha + 0.0055\alpha \sin \theta$	I
11	$\beta - 13^{\circ}12' = +0.9689\alpha + 0.2476\alpha \sin \theta$	I
12	$\beta - 13^{\circ}11' = +0.8878\alpha + 0.4602\alpha \sin \theta$	I
13	$\beta - 13^{\circ}34' = +0.7549\alpha + 0.6558\alpha \sin \theta$	III
14	$\beta - 13^{\circ}53' = +0.5755\alpha + 0.8178\alpha \sin \theta$	III
15	$\beta - 13^{\circ}58' = +0.3608\alpha + 0.9326\alpha \sin \theta$	III
16	$\beta - 14^{\circ}14' = +0.1302\alpha + 0.9915\alpha \sin \theta$	III
17	$\beta - 14^{\circ}56' = -0.1068\alpha + 0.9943\alpha \sin \theta$	III
18	$\beta - 14^{\circ}47' = -0.3363\alpha + 0.9418\alpha \sin \theta$	II
19	$\beta - 15^{\circ}56' = -0.8560\alpha + 0.5170\alpha \sin \theta$	II
20	$\beta - 13^{\circ}29' = +0.8002\alpha + 0.5997\alpha \sin \theta$	III
21	$\beta - 15^{\circ}55' = -0.9952\alpha - 0.0982\alpha \sin \theta$	II
22	$\beta - 15^{\circ}39' = -0.8409\alpha + 0.5412\alpha \sin \theta$	II
23	$\beta - 16^{\circ}9' = -0.9429\alpha + 0.3330\alpha \sin \theta$	II
24	$\beta - 16^{\circ}22' = -0.9768\alpha + 0.2141\alpha \sin \theta$	II
25	$\beta - 15^{\circ}38' = -0.6262\alpha - 0.7797\alpha \sin \theta$	II
26	$\beta - 14^{\circ}54' = -0.4091\alpha - 0.9125\alpha \sin \theta$	II
27	$\beta - 13^{\circ}7' = +0.9284\alpha - 0.3716\alpha \sin \theta$	I

Source: Mayer (1750, p. 153).

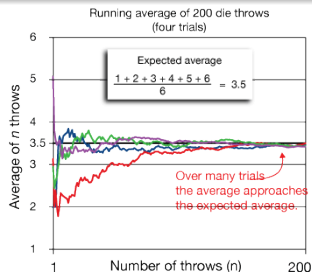
Pour la détermination de la longitude d'un navire en pleine mer.

# Statistique et probabilité

Arbutnot's (SC, LP) 1710 : test du signe pour prouver que le taux de natalité des femmes n'est pas  $1/2$ .

## La loi des grands nombres (Bernoulli, 1713)

$$\begin{array}{ccc} \text{La moyenne} & \text{converge vers} & \text{l'espérance} \\ \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i & \xrightarrow{n \rightarrow \infty} & \mathbb{E}(X) \end{array}$$



## le taux de natalité des femmes est il 1/2 ?

on a observé  $n_f = 342986$  naissances de filles sur  $n = 701819$  naissances

$$\hat{p} = \frac{n_f}{n} = 0,49$$

C'est une moyenne avec  $B_i = 1$  pour une fille et  $B_i = 0$  pour un garçon

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n B_i$$

Peut-on en déduire raisonnablement que  $p = 1/2$  ?

L'écart observé entre 0,49 et 0,5 est-il dû au hasard ou au fait que  $p \neq 1/2$

### Le théorème central limite (De Moivre, 1733)

La moyenne	converge vers	la loi normale
$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n B_i - p \right)$	$\xrightarrow[n \rightarrow \infty]{(d)}$	$\mathcal{N}(0, \sigma^2)$

$$\mathbb{P} \left( -2,57 \leq \frac{\sqrt{n}}{\sqrt{p(1-p)}} \left( \frac{1}{n} \sum_{i=1}^n B_i - p \right) \leq 2,57 \right) = 0,99$$

avec  $\left( \frac{1}{n} \sum_{i=1}^n B_i - p \right) = -0.011$  et  $\frac{2,57}{2\sqrt{701819}} = 0.0000018$



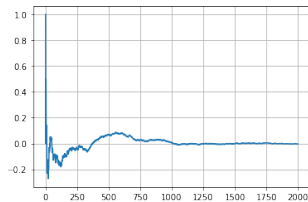
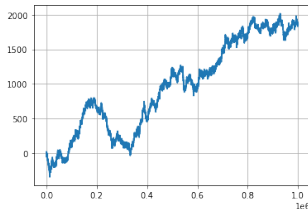
# 18ème siècle : une histoire de convergence

Soit  $X$  une variable aléatoire  $\mathbb{E}(X) = 0$  et  $V(X) = 1$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n X_i$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$$

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$$



# Physique statistique

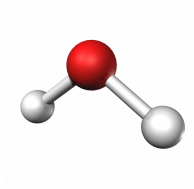
Boltzmann et Gibbs développent la mécanique statistique (1860)

Perception macroscopique  $n = 10^{23}$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}(X)$$

Ca marche !

- Pression = somme des chocs des molécules
- Déterministe mécanique  $\rightarrow$  distribution probabiliste
- Pont entre micro et macroscopique : capacité à modéliser



# de la physique statistiques vers les autres sciences

- Biologie
  - ▶ santé - médecine
  - ▶ génomique
  - ▶ ...
- Chimie
  - ▶ PSL
- dans l'industrie
  - ▶ contrôle qualité
- Sciences sociales
  - ▶ économie
  - ▶ sondages
  - ▶ langage
  - ▶ ...
- ...

# Enjeux actuels des statistiques

- Big data pour quoi faire : prédiction du jour de sa mort (Don't look up et Abraham de Moivre)
- réduire l'individu à ses propres chiffres
- les stat pour décrire le monde Robert C. Camp le père du benchmarking (Xerox). La quantification transforme le monde
- Statistiques et IA  
→ ChatGPT



BRONTEROC

Bronteroc is a four-headed four-legged creature with four legs like human hands, four legs like human legs, and a body like a unicorn.

DÉTECTER LA LANGUE

ANGLAIS

FRANÇAIS

ARABE



FRANÇAIS

ANGLAIS

ARABE

© OpenAI

La femme a arrêté de prendre la pilule parce qu'elle était enceinte.



The woman stopped taking the pill because she was pregnant.

La femme a arrêté de prendre la pilule parce qu'elle était périmée.

The woman stopped taking the pill because it had expired.

# Enjeux de toujours des statistiques

Quelles sont les données justes pour la question que je me pose ?

Produire de la connaissance à partir de données est un problème mal posé.

Pour le résoudre il faut des a = **Un modèle statistique**

Statistique = probabilité<sup>-1</sup>

Proba

Inférence

# Objectifs du cours de statistiques pour l'ingénieur

## Objectifs

- introduire les principales méthodes statistiques utilisées en ingénierie ;
- donner une introduction à la théorie mathématique derrière les méthodes statistiques ;
- donner des garanties théoriques sur les méthodes statistiques que vous pourrez utiliser pour certaines applications.

A la fin de ce cours, vous devriez être en mesure de :

- à partir d'une situation réelle, formuler un problème statistique en termes mathématiques
- sélectionner les méthodes statistiques appropriées à votre problème
- comprendre les implications et les limites des différentes méthodes