

Quelques Repères Historiques : Les Statistiques

La statistique ne se limite pas à la récolte et l'analyse de données. Elle est une branche des mathématiques qui utilise les probabilités pour prévoir à partir de données réelles. Son champ d'application concerne toutes les sciences.

Sondages, Enquêtes, ...

Déjà les anciens empires de Chine (XXIII^e av. J.-C.) et d'Egypte (XVIII^e av. J.-C.) recensent leur production agricole et collectent des données. Ces activités correspondent aux besoins des Etats ou des empires de quantifier les ressources en hommes et en biens et de répartir les charges et impôts.

Les premiers progrès sont dus à l'utilisation de tableaux et de représentations graphiques mais jusqu'à la fin du XVII^e siècle, la statistique reste descriptive.

Le terme *statistiques* apparaît pour désigner les enquêtes sur les ressources nationales au XVIII^e siècle.

Status en latin : état.

Un outil de gouvernement

Le royaume de France cherche, au XVIII^e siècle à connaître sa population. Le mathématicien **Laplace** calcule le nombre de paroisses à interroger pour étendre ses résultats à la France entière avec une erreur inférieure à 5%. La question de la représentativité ainsi posée est au cœur de nos sondages modernes.



Pierre Simon Laplace 1749-1827

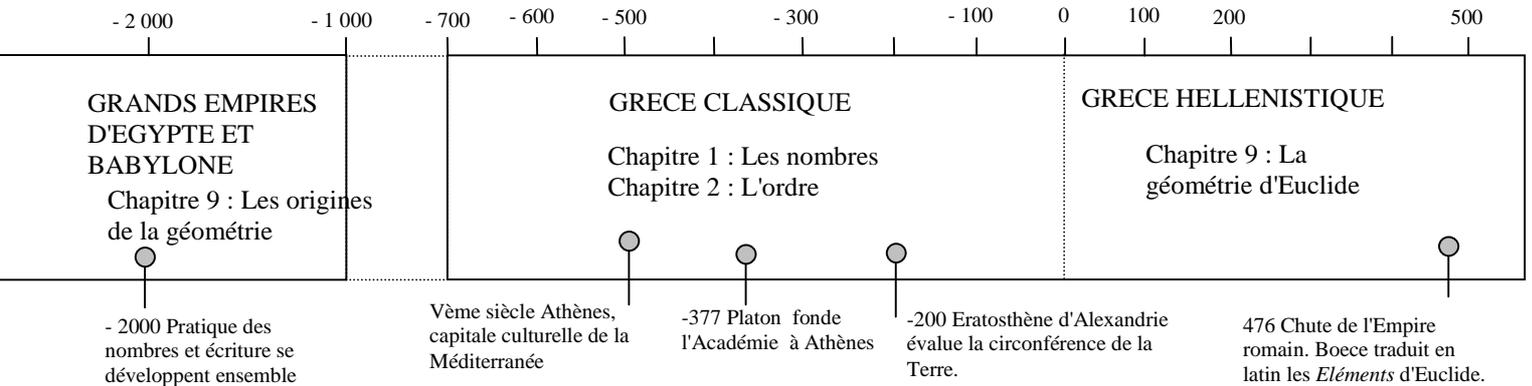
Huygens et l'espérance de vie

Si l'idée de collecter des données remonte à l'Egypte ou la Chine ancienne, les premières tables statistiques pour le commerce, les assurances, la démographie, c'est-à-dire l'économie et les sciences sociales en général datent du XVII^e. On en a un exemple avec Christian Huygens (1629-1695), physicien et mathématicien hollandais. Auteur du premier ouvrage sur le calcul des probabilités en 1657, il y introduit explicitement la notion d'espérance. Il s'est également beaucoup intéressé aux statistiques démographiques. Les outils de la statistique moderne ne naissent

que plus tard : *moyenne* et *médiane* au XVIII^e siècle, *variance* et *écart type* au XIX^e.

La moyenne

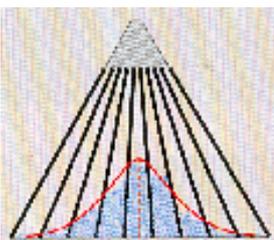
Pendant longtemps, on a considéré la moyenne arithmétique comme un paramètre pouvant traduire l'ensemble des données observables. Mais les valeurs trouvées sont sujettes à des erreurs et à débats. Diverses méthodes ont été mises au point par Gauss et Legendre (méthode des moindres carrés) et Laplace (loi normale). Au XIX^e siècle, la synthèse de tous les travaux a été réalisée par l'astronome et statisticien belge **Adolphe Quételet**. Ses travaux le conduisent à définir la notion d'"homme moyen" dont chaque caractéristique serait constituée de la moyenne de toutes les mesures de tous les hommes. Les différents calculs et méthodes mènent donc à des interprétations contestables ou abstraites.



Interpréter les données

Comment résumer les nombreuses observations menées sur une étoile ?

La théorie développée par Gauss fait de la mesure moyenne la meilleure estimation de la grandeur cherchée.



Courbe de Gauss : Répartition autour de la moyenne



Carl Friedrich Gauss 1777-1855

L'analyse des fréquences

Un des moyens les plus simples de chiffrer un message est de remplacer chaque lettre par une autre (ou un autre symbole). Par sa simplicité et par sa force, ce système a dominé la technique des écritures secrètes pendant tout le premier millénaire. Il a résisté aux cryptanalystes jusqu'à ce que le savant arabe Abu Yusuf Ya'qub ibn Is-haq ibn as-Sabbah Oòmram ibn Ismaïl **Al-Kindi** mette au point, au IX^e siècle, une technique appelée analyse des fréquences.

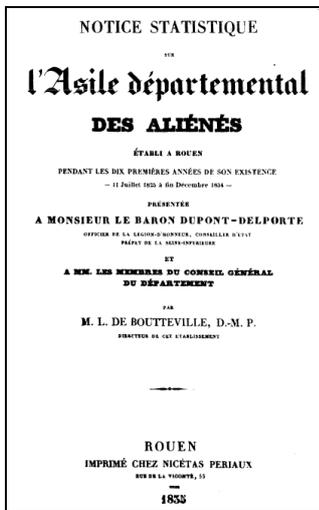
Une analyse semblable sur les fréquences d'apparition des chiffres dans un ensemble de données est modélisée par la loi de Benford



L'asile pour aliénés de Rouen :

Un laboratoire de statistiques morales de la Restauration à 1848

De 1825 à 1848, l'asile Saint-Yon de Rouen pour aliénés est un véritable « laboratoire de statistiques sociales et morales ». En 1835, le docteur Lucien Deboutteville fait paraître une première notice statistique.



Elle fournit aux autorités municipales de nombreuses informations sur le nombre, l'âge et le sexe de la population asilaire. Elle renseigne aussi sur le taux de décès à l'intérieur de l'asile et l'influence des saisons sur la maladie mentale.

Une définition de la statistique

Nous entendons par Statistique, la science qui a pour objet de recueillir et de coordonner des faits nombreux sensiblement indépendants des anomalies du hasard, et qui dénotent l'existence des causes régulières dont l'action s'est combinée avec celle des causes fortuites, par Cournot (1801-1877) dans *Exposition de la théorie des chances et des probabilités* (1843)

Le hasard

Hasard et aléas viennent tous deux de mots arabes et latins signifiant "jeux de dés". Dans l'antiquité, le hasard est un instrument de divination et une aide à la prise de décision. Les régularités observées, lors d'expériences répétées, ont permis progressivement de se détacher de ces références liées au destin.



Blaise Pascal 1623-1662

Des échanges épistolaires entre Pascal et le chevalier de Méré, grand amateur de jeux de hasard au XVII^e siècle naît une théorie du hasard puis des probabilités.

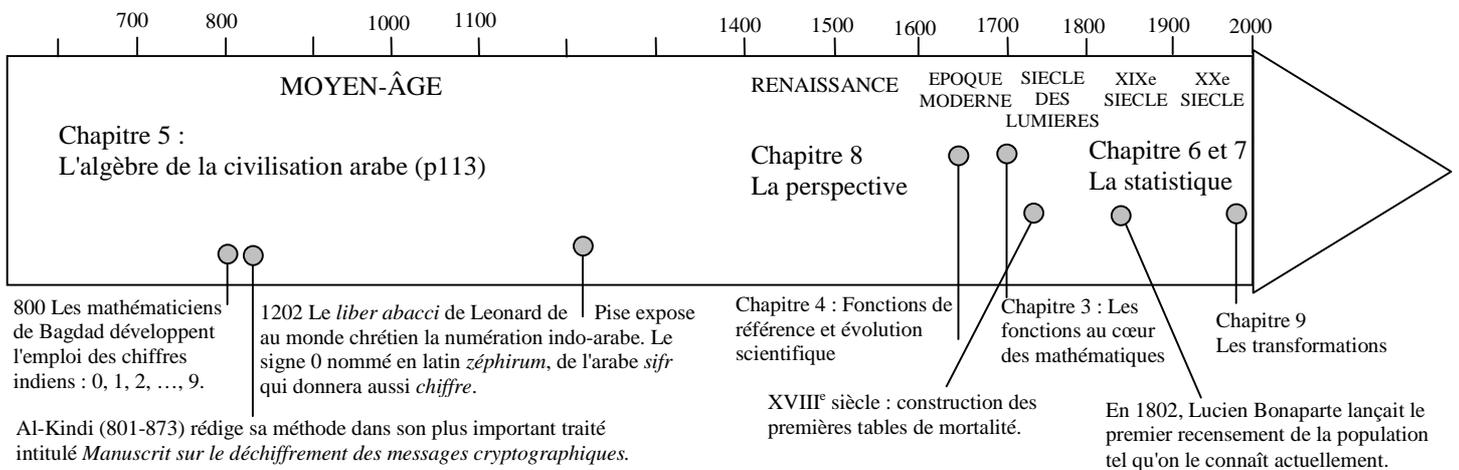
Domestiquer le hasard

Etat, entreprises, instituts scientifiques, financent des collectes de données (l'INSEE est créée en 1946) gérés par des ordinateurs et diffusés par des médias.



Institut National de la Statistique et des Études Économiques

Taux de croissance, sondages, simulations météorologiques : la statistique moderne donne forme à la réalité et sait, paradoxalement, utiliser le hasard pour rendre le monde moins capricieux, moins imprévisible.



Quelques grands noms

Ronald Aylmer Fisher (1890-1962) : on a pu dire de lui qu'il est "l'homme qui a fait de la statistique une science moderne".



William Gosset (1876-1937) : Il a contribué à développer les statistiques industrielles. Employé par les Brasseries Guinness, il a du utiliser le pseudonyme de *Student* afin de publier ses travaux liés à la statistique. La loi dite de *Student* est utilisée dans des estimations basées sur de petits échantillons.

Jerzy Neyman (1894-1981) : il est le créateur de la théorie moderne des sondages.

Abraham Wald (1902-1950) : il a élaboré la théorie des décisions statistiques. Il a développé, entre autre, une méthode d'estimation de la vulnérabilité des avions. En 1947, certains de ses travaux furent classés "secret défense" par le gouvernement américain.

Un outil de preuve scientifique :

Les diagrammes inventés par Florence Nightingale démontrent la terrible conséquence des conditions d'hygiène sur la mortalité. La statistique donne à la scientifique britannique l'autorité pour initier des réformes sanitaires dans toute l'Angleterre.

La statistique est un outil de preuve scientifique qui, au XX^e siècle, a envahi tous les domaines du savoir.

