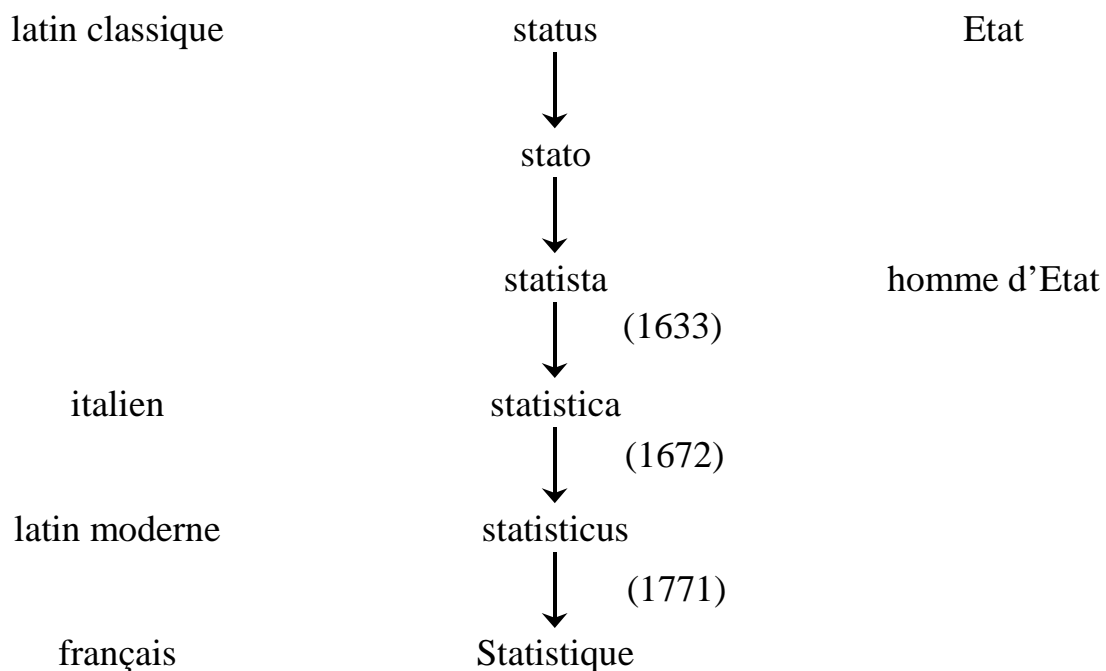


Chapitre 1:

Introduction

1.1. Origine

L'origine du mot « statistique » remonte au latin classique *status* (état) qui, par une série d'évolutions successives, aboutit au français *statistique*, attesté pour la première fois en 1771.



C'est vers la même époque que *statistik* apparaît en allemand, alors que les anglophones utilisent l'expression *political arithmetic* jusqu'en 1798, date à laquelle le mot *statistics* fait son entrée dans cette langue.

1.2. Statistique et statistiques

A l'origine, cette discipline concerne donc les affaires de l'Etat.

Actuellement, on distingue généralement *les statistiques* (au pluriel) de *la statistique* (au singulier).

Les statistiques peuvent être définie comme l'étude méthodique des faits sociaux qui définissent un Etat, par des procédés numériques (dénombrements, inventaires, recensements,...)

Parmi les premiers ouvrages ayant trait à cet aspect, on peut mentionner « Le Détail de la France », rédigé par Boisguilbert en 1697 et, en 1722, « Description de la France », dû à Pigonal de la Force.

Le second sens n'apparaît que vers 1830. C'est celui qui est abordé dans ce cours. Nous définirons *la statistique* comme un **ensemble de techniques d'interprétation mathématique appliquées à des phénomènes (ex : faits sociaux) pour lesquels une étude exhaustive de tous les facteurs est impossible à cause de leur grand nombre ou de leur complexité.**

1.3. Historique

De tous temps, les chefs d'Etat ont souhaité déterminer la puissance des nations qu'ils dirigeaient à l'aide de recensements partiels ou complets (population, territoire, production,...)

Dès 3000 av. J.-C., on trouve mention de collectes d'observations sur les biens et les personnes en Mésopotamie.

En 1200 av. J.-C., des évaluations de productions agricoles sont effectuées en Chine.

Au début de notre Ere a lieu un dénombrement des richesses de l'Empire Romain, rendu célèbre par sa mention dans l'Evangile de Luc.

Au Moyen Age, des relevés sont exécutés sur l'ordre de Charlemagne puis de Guillaume le Conquérant. Dans les deux cas, le but est de se faire une idée plus précise des richesses du pays.

Au XVII^e Siècle, pour éviter le recensement lourd et onéreux, William Petty (1623-1687) met au point une méthode de comptage de la population de Londres sur base des proportions moyennes entre :

- les maisons
- les feux (ménages) par maison
- la composition des familles

Au XIX^e Siècle, les recensements proprement dits reprennent de l'importance et, en 1853, a lieu à Bruxelles le 1^{er} Congrès International de Statistique, sous l'impulsion d'Adolphe Quételet (1796-1874, astronome et mathématicien belge, un des fondateurs de la science statistique). L'objectif de ce congrès est d'uniformiser les techniques de compilation des statistiques nationales, en vue de faciliter les comparaisons.

Au début du XX^e Siècle, un débat oppose les partisans des recensements (réalisés sur l'ensemble de la population) et des sondages (réalisés sur un échantillon représentatif de la population).

Les recensements ne sont pas toujours possibles, ni souhaitables. Dans certains cas, ils peuvent être trop chers (comme, par exemple, des enquêtes sur toute la population d'un pays). Ils peuvent aussi contenir des erreurs. Parfois, ils sont carrément aberrants (mesurer la solidité moyenne d'un type de voiture en lançant toutes les voitures de ce type contre un mur serait commercialement inacceptable).

Pour pallier ces inconvénients, on a recours au sondage statistique, qui consiste à

déduire les propriétés de toute une population à partir de l'analyse d'un échantillon

Il est capital que l'échantillon soit choisi et analysé de manière adéquate. En particulier, il faut que l'échantillon soit *représentatif* de la population. Un échantillon non représentatif est dit *biaisé*.

Au début du XX^e siècle, beaucoup de journaux américains réalisent des « votes de paille » en demandant leur avis par écrit à plusieurs millions de personnes quelques semaines avant les élections présidentielles.

En 1936, le Literary Digest prédit, à l'aide d'un échantillon de 2.400.000 électeurs, la victoire du candidat républicain N. Landon. George Gallup, grâce à un sondage sur 4000 personnes judicieusement choisies, prévoit, quand à lui, la victoire du démocrate Franklin D. Roosevelt. La victoire de ce dernier sonne le glas des votes de paille dont les échantillons sont souvent biaisés (les cartes du Literary Digest avaient été envoyées aux abonnés du téléphone et aux propriétaires de voitures, cet électorat aisé était plus favorable aux républicains).