

Statistique

Introduction

Angelina Roche
roche@ceremade.dauphine.fr

Université Paris-Dauphine PSL

15 septembre 2022

Retour en arrière : les données hier...

- ▶ « Statistik » (dérivé du latin *Statisticum*). Allemagne, 1740 (Achenwall). **Ensemble de mesures** et recueil de données nécessaires au fonctionnement et à l'organisation de l'état : recensements et estimations de la population, des richesses, de l'impôt, des armées.
- ▶ Les progrès de la statistique : représentation graphique et **organisation des données en tableaux** (statistique descriptive, dominée par l'école allemande au 18^e siècle). Activité importante aussi en Grande Bretagne¹ et dans une France centralisée.²

1. William Playfair (1759–1823), 1786, "The Commercial and Political Atlas" contenant le premier diagramme en barres connu.

2. Vauban, 1686, "Méthode générale et facile pour faire le dénombrement des peuples".

Les données aujourd'hui : (1) chiffres du travail

Les chiffres du travail

Taux d'activité par tranche d'âge hommes vs. femmes

	A	B	C	D	E	F	G	H	I
1									
2	Taux d'activité par tranche d'âge de 1975 à 2005								
3	En %								
4		1975	1976	1977	1978	1979	1980	1981	1982
5	Femmes								
6	15-24 ans	45,5	45,7	45,2	43,9	44,2	42,9	42,1	41,87
7	25-49 ans	58,8	60,3	62,1	62,8	64,7	65,4	66,2	67,55
8	50 ans et plus	42,9	43,1	44,4	43,9	44,8	45,9	45,2	43,47
9	Ensemble	51,5	52,5	53,6	53,6	54,8	55,1	55,1	55,29
10	Hommes								
11	15-24 ans	55,6	54,7	53,7	52,2	52,5	52,0	50,4	45,02
12	25-49 ans	97,0	97,1	96,9	96,9	96,9	97,1	96,9	96,75
13	50 ans et plus	79,5	78,8	79,5	78,8	79,4	78,3	75,4	71,65
14	Ensemble	82,5	82,2	82,1	81,6	81,8	81,5	80,4	78,14

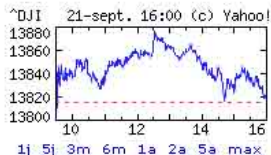
<http://www.insee.fr/>

Les données aujourd'hui : (2)

donnés financières

DOW JONES INDUSTRIAL AVERAGE IN (DJI; ^DJI)

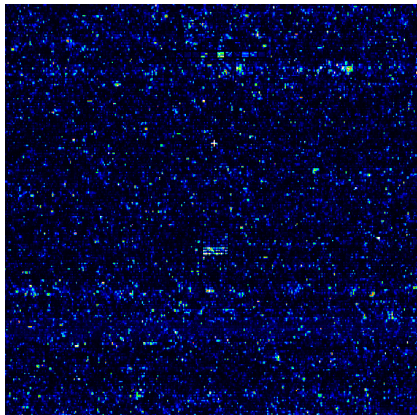
Dem. Cours:	13.820,19
Heure:	21 sept.
Variation:	↑ 53,49 (0,39%)
Clôture Préc.:	13.766,70
Ouverture:	13.768,33
Var. Journalière:	13.768,25 - 13.877,17
Var. sur 1 an:	11.926,80 - 14.121,00
Volume:	419.389.397



<http://fr.finance.yahoo.com/>

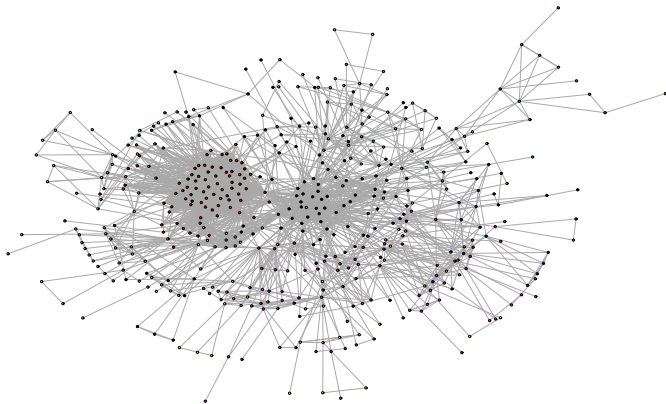
Les données aujourd'hui : (3)

Biopuces et analyse d'ADN




Les données aujourd'hui : (4)

E-marketing - Livres



Statistique et probabilités

- ▶ 17^e siècle : **Invention des probabilités**. Incorporation d'un raisonnement probabiliste – et donc un modèle du hasard – dans le traitement d'observations.
- ▶ Basculement de la statistique vers une **discipline scientifique** à part entière. Préfigure l'actuariat moderne³.
- ▶ L'exemple historique incontournable : **John Arbuthnott** (1667–1735) et le déficit des naissances et morts selon le sexe. – la première réflexion « moderne » en statistique.

3. les frères Hyugens, premier calcul de l'espérance de vie humaine en 1669, Graunt (1620–1674), William Petty (1623–1687), Laplace. 

John Arbuthnott et « la divine providence »

- ▶ 1712, Arbuthnott (médecin de la Reine Anne) examine le nombre de baptêmes de filles et de garçons à Londres, entre 1629 et 1710.
- ▶ Sur 82 années retenues, le nombre de naissances masculines est toujours supérieur au nombre de naissances féminines.
- ▶ Arbuthnott calcule la probabilité que les naissances masculines (avec équiprobabilité de l'apparition fille/garçon) soient plus nombreuses que les naissances féminines, 82 fois de suite
- ▶ Cette probabilité = $(1/2)^{82}$, « *which will be found easily by the Table of Logarithms to be 1/4 8360 0000 0000 0000 0000 0000* ».

An Argument for Divine Providence

- ▶ *An Argument for Divine Providence, taken from the constant Regularity observed in the Births of both Sexes*
- ▶ « [...] *This Event is wisely prevented by the Oeconomy of Nature; and to the judge of the wisdom of the Contrivance, we must observe that the external Accidents to which Males are subject (who must seek their food with danger) do make a great havock of them, and that this loss exceeds far that of the other Sex, occasioned by Disease incident to it, as Experience convinces us. To repair that Loss, provident Nature, by the Disposal of its wife Creator, brings more Males than Females; and this in almost a constant proportion* » .

Raisonnement statistique

- ▶ Point de départ : des observations

$$\mathbf{x}_1, \dots, \mathbf{x}_n$$

et une **problématique** (estimer la moyenne, déterminer si la proportion dépasse un certain seuil,...).

- ▶ **Modélisation statistique**

1. Modéliser $\mathbf{x}_1, \dots, \mathbf{x}_n$ comme les réalisations

$$X_1(\omega), \dots, X_n(\omega)$$

de variables aléatoires X_1, \dots, X_n .

2. La **loi** de (X_1, \dots, X_n) **est inconnue**.

- ▶ Comment répondre à la problématique initiale à partir de nos observations ?
- ▶ Quelle **incertitude** sur le résultat obtenu ?

Retour sur le problème d'Arbuthnott

► Les observations :

$x_i = 1$ si le nombre de naissance de garçons est supérieur à celui de filles l'année i ($i \in \{1, \dots, n\}$, $n = 82$),

$x_i = 0$ sinon.

Il observe $(x_1, \dots, x_n) = (1, \dots, 1)$.

► Problématique : les données observées sont-elles liées à de simples fluctuations aléatoires ou la probabilité d'une naissance masculine est-elle supérieure à $1/2$?

► Modélisation statistique

1. Modéliser x_1, \dots, x_n comme les réalisations

$$X_1(\omega), \dots, X_n(\omega)$$

de variables aléatoires $X_1, \dots, X_n \sim_{i.i.d} \mathcal{B}(p)$.

2. Si l'on fait l'hypothèse que $p = 1/2$, alors

$$\mathbb{P}(X_1 = 1, \dots, X_n = 1) = \mathbb{P}(X_1 = 1) \dots \mathbb{P}(X_n = 1) = p^{82} = (1/2)^{82}.$$

C'est un des premier **test statistique** !

Plan du cours

1. Rappels de probabilité

comprendre et poser les modèles probabilistes.

2. Intervalles de confiance

mesurer les fluctuations.

3. Construction d'estimateurs

estimer des paramètres autres que la moyenne et la proportion

4. Quelques tests statistique courants.

Evaluation

DST 1 jeudi 20 octobre de 10h30 à 12h30,

séance de questions/réponses facultative de 9h30 à 10h30,

polycopié de cours et calculatrices autorisés.

DST 2 jeudi 1er décembre de 10h30 à 12h30,

séance de questions/réponses facultative de 9h30 à 10h30,

polycopié de cours et calculatrices autorisés.

Projet (facultatif) à rendre avant le **jeudi 10 novembre**. Plus d'information sur ceremade.dauphine.fr/~roche/enseignement

Note finale

$\max\{0.5 DST1 + 0.5 DST2; 0.4 DST1 + 0.4 DST2 + 0.2 projet\}$.