

# Statistiques et Informatique Appliquées à la Sociologie

## UV SOC23A1

### Présentation du cours 2001/2002

#### Organisation matérielle

Cours magistral : 7 ou 8 heures

mercredi 13h45-14h45 - A222

Travaux dirigés de statistiques : 1 h par quinzaine

– mardi 8h15 à 9h15 - sem A - Gr. 1 - A221

– mardi 8h15 à 9h15 - sem B - Gr. 2 - A221

Travaux dirigés d'informatique : 2 h par quinzaine. Effectif max. : 20

– mercredi 8h15-10h15 - sem A - A203

– mercredi 8h15-10h15 - sem B - A203

– mercredi 13h45-15h55 - sem A - A203

– mercredi 13h45-15h55 - sem B - A203

– vendredi 8h15-10h15 - sem A - A203

Contrôle des connaissances : (contrôle continu)

Examen écrit (1 heure) en janvier

Un dossier

## **Bibliographie**

- Bloss et Grosseti, Introduction aux méthodes statistiques en sociologie, PUF, Coll. Le Sociologue
- G. Mialaret. Statistiques appliquées aux sciences humaines. PUF
- Colin, Lavoie, Delisle. Initiation aux méthodes quantitatives en Sc. Humaines CDR. No 310-LAV V2098/A
- P. Rateau, Méthode et statistique expérimentales en sciences humaines, Ellipses

**Contenu** Pour le premier semestre : CM essentiellement consacré aux statistiques.

Statistiques descriptives : résumer, donner une vue synthétique d'un ensemble de données : représentations graphiques, paramètres de position et de dispersion  
Statistiques "mathématiques" : notion de distribution théorique.

Documents fournis :

Transparents du CM et photocopiés de TD

Egalement accessibles sur Internet (dans le courant du semestre) (au format .pdf lisible par Acrobat Reader) :

<http://geai.univ-brest.fr/~carpentier/>

### **Pourquoi faut-il étudier les statistiques ?**

Les statistiques sont-elles utiles au sociologue ?

Les statistiques, il y a des calculatrices et des logiciels pour faire cela. Oui, mais ...

# **Introduction – Vocabulaire**

## **Collecte des données**

### **Sur qui?**

Population

Individu statistique, unité statistique, sujet

### **A propos de quoi?**

Attribut, caractère, variable statistique

Modalités d'une variable: exhaustives - exclusives

Champ ou domaine de variation

## **Nature d'une variable statistique**

Variables nominales – échelle nominale

Variables ordinales – échelle ordinale

Variables numériques discrètes ou continues

Echelles d'intervalles, échelles de rapports

## Recueil et présentation d'un ensemble de données

Individus :  $s_1, s_2, \dots, s_N$

Variables étudiées :  $X, Y, \dots$

**Tableau protocole :**

	Variable $X$	Variable $Y$
Individu 1	$x_1$	$y_1$
Individu 2	$x_2$	$y_2$
...	...	...
Individu $N$	$x_N$	$y_N$

Recensement ou tri à plat : **tableau d'effectifs**

Modalités	Effectifs	Fréquences
Modalité $a_1$	$n_1$	$f_1$
Modalité $a_2$	$n_2$	$f_2$
...	...	...
Modalité $a_k$	$n_k$	$f_k$
	$N$	1 (ou 100%)

Fréquences :

$$f_i = \frac{\text{effectif de la modalité } a_i}{\text{effectif total}}$$

Variable regroupée en classes :

Classes	Effectifs	Fréquences
$[a_1, a_2[$	$n_1$	$f_1$
$[a_2, a_3[$	$n_2$	$f_2$
...	...	...
$[a_k, a_{k+1}[$	$n_k$	$f_k$
	$N$	1 (ou 100%)

## Tableau de données chronologiques

Exemple : étude d'une même variable à deux moments différents.

CSP	1962	1990	Ec. abs.	Ec. rel.	Coeff.
Agriculteurs	11	38	27	245 %	3.45
Sal. agricoles	1	5	4	400 %	5
Patrons	37	75	38	103 %	2.03
...	...	...	...	...	...
Total	197	894	697	354 %	4.54

*Ecart absolu* entre deux effectifs d'une même modalité.

*Ecart relatif* ou taux de variation :

$$\text{Ecart relatif} = \frac{\text{Effectif 1990} - \text{Effectif 1962}}{\text{Effectif 1962}}$$

*Coefficient multiplicateur*:  $1 + \text{Ecart relatif}$

*Taux de variation annuel moyen* entre 1962 et 1990 :

$$\text{Coeff. moyen} = 3.45^{\frac{1}{28}} = 1.045$$

$$\text{Taux moyen} = 4.5\%$$

Etude conjointe de deux variables ou tri croisé: **ta-  
bleau de contingence**

$X \setminus Y$	$b_1$	$b_2$	...	$b_j$	...	$b_l$	
$a_1$	$n_{11}$	$n_{12}$		$n_{1j}$	...	$n_{1l}$	
$a_2$	$n_{21}$	$n_{22}$			...		
...							
$a_i$	$n_{i1}$	$n_{i2}$		$n_{ij}$	...		
...							
$a_k$	$n_{k1}$	$n_{k2}$			...	$n_{kl}$	
	$N_{.1}$						$N$

**Exemple**

	Hommes	Femmes	Total
Comédie	90	150	240
Drame	50	90	140
Variétés	160	160	320
	300	400	700

Deux (voire trois) manières de calculer des fréquences :

Par ligne

	Hommes	Femmes	Total
Comédie	37.5%	62.5%	100%
Drame	35.7%	64.3%	100%
Variétés	50%	50%	100%
	42.8%	57.2%	100%

Par colonne

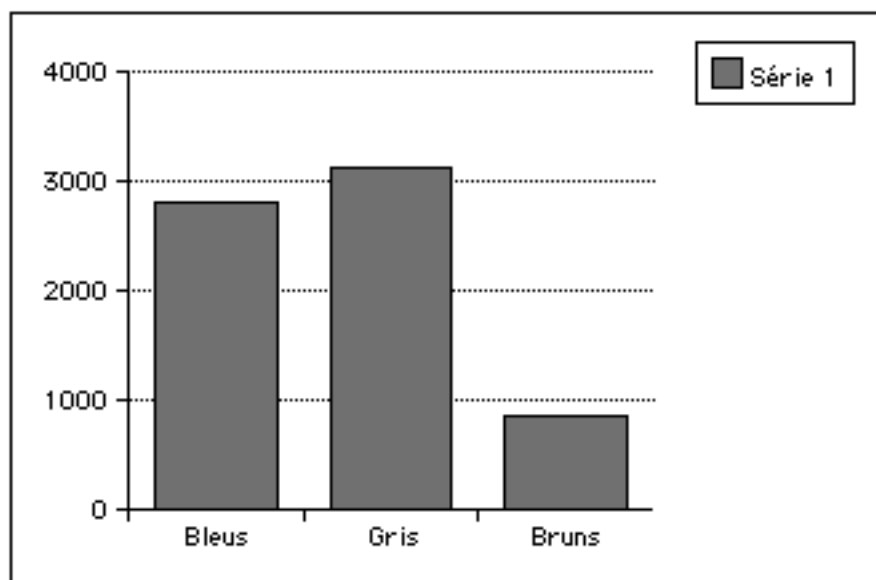
	Hommes	Femmes	Total
Comédie	30%	37.5%	34.3%
Drame	17%	22.5%	20%
Variétés	53%	40%	45.7%
	100%	100%	100%

# Représentations graphiques

## Variables nominales

Mod.	Eff.	Freq.	%
Bleus	2811	0,41	41%
Gris	3132	0,46	46%
Bruns	857	0,13	13%
Total	6800	1	100%

## Diagrammes à bandes

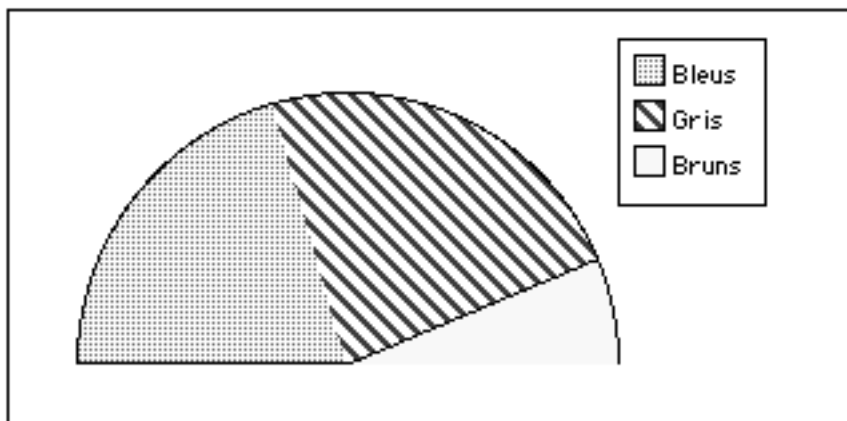




## Diagrammes circulaires ou semi-circulaires

Méthode de construction : construire un tableau de proportionnalité

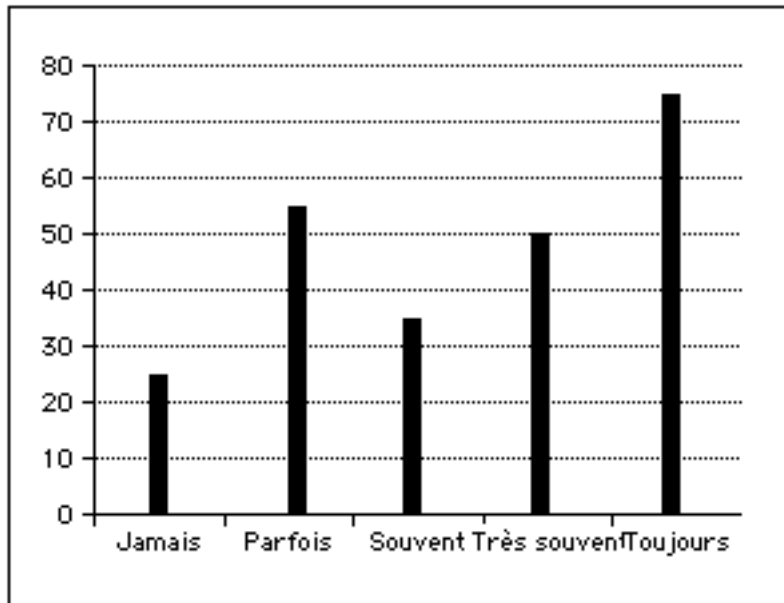
Modalités	Bleu	Gris	Bruns
Fréq.	0,41	0,46	0,13
Angle	74	83	23



## Variable ordinale : diagramme en bâtons

On a demandé à 240 sujets s'ils fermaient à clef la porte de leur appartement.

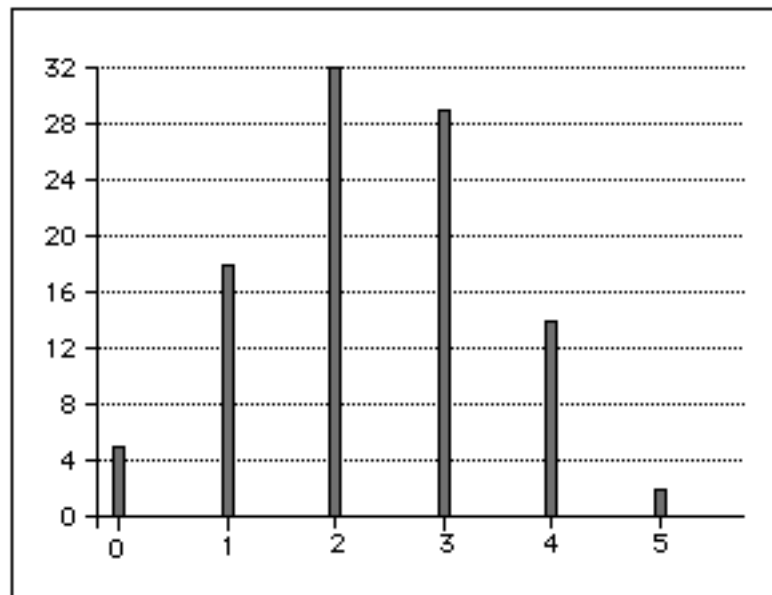
	Jam.	Parf.	Souv.	T. souv.	Tjrs
Eff.	25	55	35	50	75



**Variable numérique discrète** : même type de construction ; graduation régulière sur l'axe des abscisses.

Exemple : nombre de garçons dans des familles de 5 enfants

Mod.	0	1	2	3	4	5
Eff.	5	18	32	29	14	2



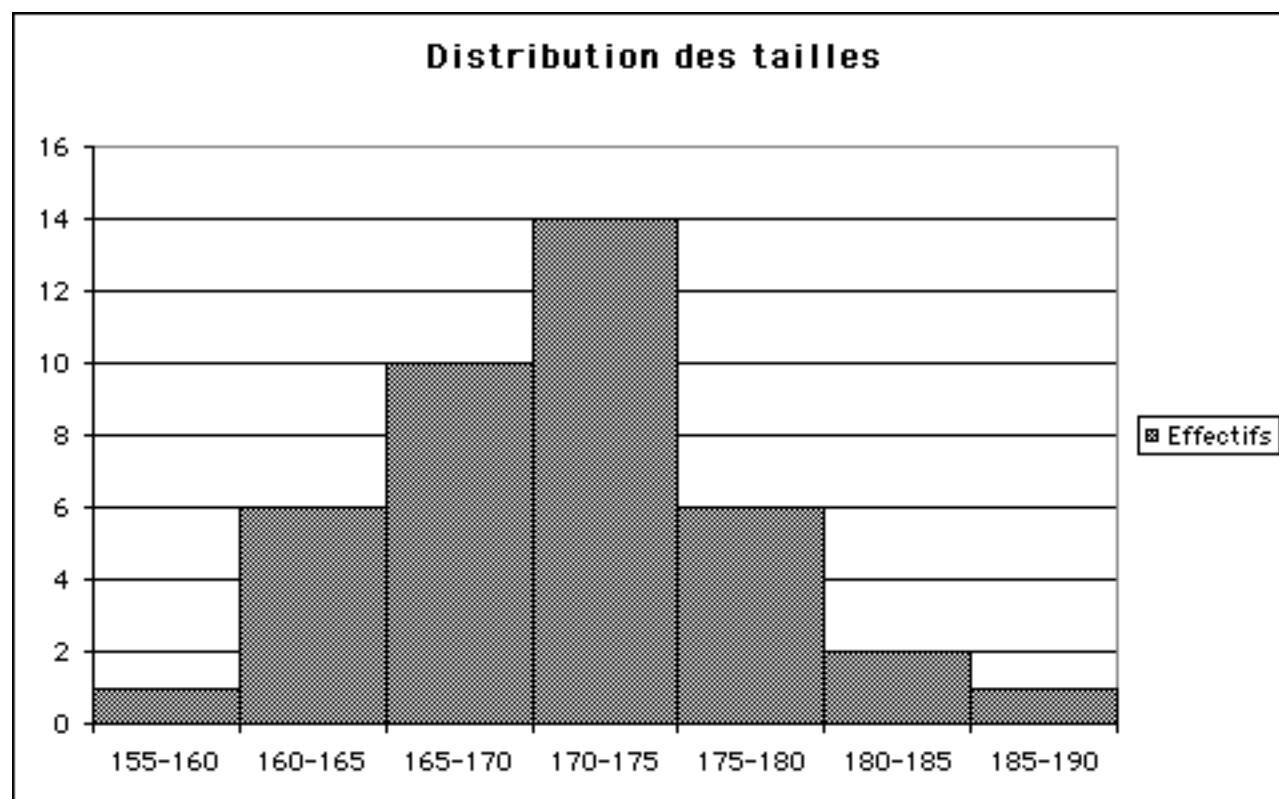
**Variable numérique regroupée en classes : histogramme.**

Un histogramme est formé de rectangles adjacents :  
– dont la base est proportionnelle à l'amplitude de la classe  
– dont l'aire est proportionnelle à l'effectif de la classe

## Classes de même amplitude

Classe	Eff.
[155, 160[	1
[160, 165[	6
[165, 170[	10
[170, 175[	14
[175, 180[	6
[180, 185[	2
[185, 190[	1

Hauteur des rectangles : effectifs des classes

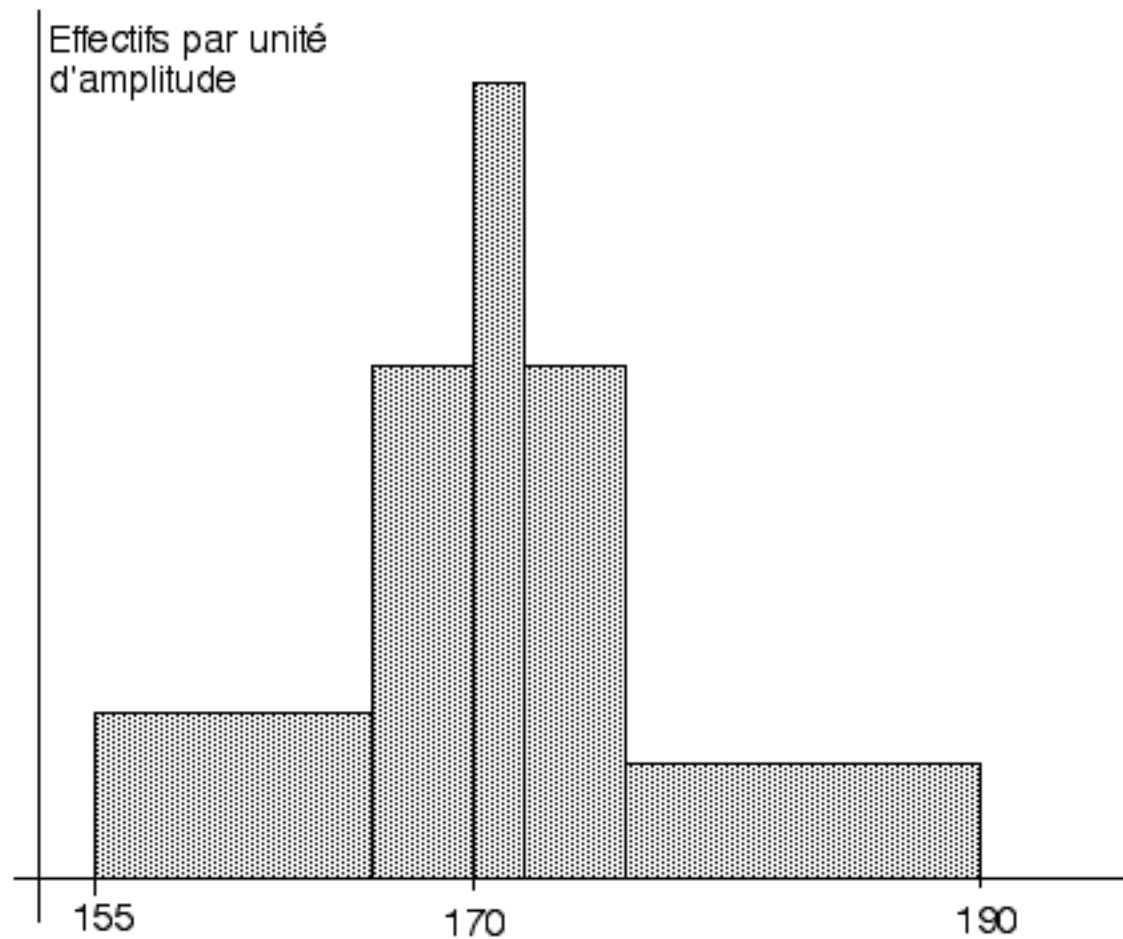


## Classes d'amplitudes différentes

$$\text{densité} = \frac{\text{effectif}}{\text{amplitude}}$$

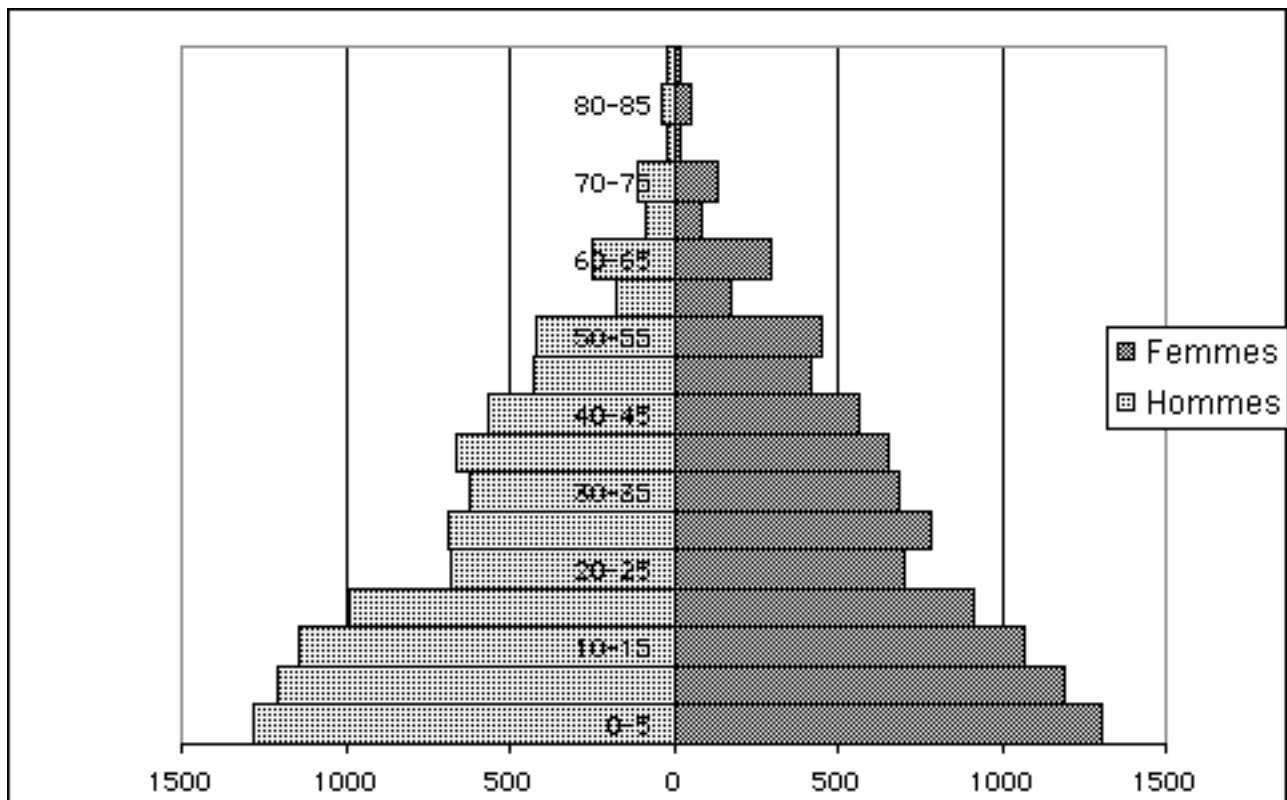
Hauteur des rectangles : densités des classes

Classe	Eff.	Amplitude	Densité
[155, 166[	8	11	0.73
[166, 170[	9	4	2.25
[170, 172[	7	2	3.5
[172, 176[	9	4	2.25
[176, 190]	7	14	0.5



## Représentation de plusieurs variables sur un même graphique

Exemple : pyramide des âges



# Caractéristiques de position

## Mode, classe modale

Mode d'une série statistique (nominale, ordinale ou numérique) : modalité correspondant à l'effectif le plus élevé.

N.B. Une série statistique peut admettre plusieurs modes.

Classe modale d'une série statistique regroupée en classes : classe qui a la plus forte densité.

N.B. : c'est la classe correspondant au rectangle de hauteur maximale dans l'histogramme.

## Médiane

Variable ordinale ou numérique.

Les individus sont *classés par valeurs croissantes de la variable*. La médiane est la valeur du caractère observée sur l'individu "médian", à savoir :

– Si  $N$  est impair, la médiane est la modalité observée sur l'individu de rang  $\frac{N+1}{2}$

– Si  $N$  est pair et si le caractère est numérique, la médiane est la moyenne des modalités observées sur les individus de rangs  $\frac{N}{2}$  et  $\frac{N}{2} + 1$ .

## Moyenne arithmétique

Caractère numérique.

– Calcul à partir d'un tableau protocole

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

– Calcul à partir d'un tableau d'effectifs

$$\mu = \frac{1}{N} \sum_{i=1}^k n_i a_i = \sum_{i=1}^k f_i a_i$$