

S. IRAQUI
Enseignant HEM

Mars 20

LA STATISTIQUE

Discipline regroupant des techniques et méthodes permettant de

- Collecter des données
- Les organiser (Décrire - Résumer)
- Modéliser et expliquer leur évolution et/ou leur variabilité

Dans le but

- d'informer
- d'aider à la prise de décision

DEUX PARTIES

- **Statistique descriptive**
 - exploration des données
 - mesures et indices
 - représentations
graphiques
- **Statistique inférentielle**
 - tester des hypothèses
 - faire des prédictions à
partir d'échantillons

EXEMPLES

Lorsque

- je calcule la moyenne d'une classe
- je trace la courbe d'évolution du chômage en 2013

je fais de la *statistique descriptive*

Lorsque

- je me demande si l'effet d'un médicament est réel ou s'il est imputable au hasard
- j'essaie de prédire des intentions de vote à partir d'un échantillon de 1000 personnes

je fais de la *statistique inférentielle*

A DISTINGUER

- Statistique descriptive, étude d'un événement réalisé,
- Probabilités, étude d'un événement à réaliser,
- La statistique inférentielle, exploitation des deux techniques précédentes.

STATISTIQUE DESCRIPTIVE

- La statistique descriptive a pour but de décrire, de synthétiser et de comparer des données recueillies en fonction du lieu ou du temps.
- Elle correspond à un besoin constant de connaissances économiques, politiques et maintenant personnelles.
- C'est une nécessité de prévision et d'aide à la décision pour l'état, et pour l'entreprise.
- Après observation de faits, elle résumera un ensemble de données obtenues par des tableaux, des graphiques, des indicateurs.

PROCÉDURE STATISTIQUE

- Recueillir des données brutes
- Effectuer un rangement des données brutes.
- Présentation des résultats dans des tableaux.
- Faire une représentation graphique des données
- Calculer les caractéristiques ou indicateurs
- Analyser les caractéristiques étudiées

RECHERCHE DE L'INFORMATION

Lors d'une étude statistique on peut procéder à:

- un **recensement** : étude de la population entière
- ou à un **sondage** : choix d'une partie de la population.

MÉTHODES DE SONDAGE

- Méthode probabiliste ou aléatoire,
- la méthode des quotas,
- la méthode des strates.

LA RECHERCHE D'INFORMATION

- Peut être directe, les données sont disponibles sous forme de flux ou nécessite une recherche .

Il existe deux types de flux :

- Le flux continu : les informations sont enregistrées d'une manière continue dans le temps. (naissance, mariage, accident, divorce, décès,...)
- Le flux périodique : les informations ne sont disponibles qu'à la fin de chaque période.

QUESTIONNAIRE

Certaines informations ne sont pas disponibles et doivent être recherchées au moyen d'un questionnaire.

- Publipostage,
- Phoning,
- Mailing,
- Méthode directe.

Population statistique :

C'est l'ensemble sur lequel porte l'étude.

Echantillon:

C'est un sous ensemble de la population.

Individu (ou unité statistique) :

Les individus ou observation sont les éléments de la population statistique étudiée.

Caractère statistique ou variable statistique :

C'est ce qui est observé ou mesuré sur les individus d'une population statistique.

Modalité d'une variable:

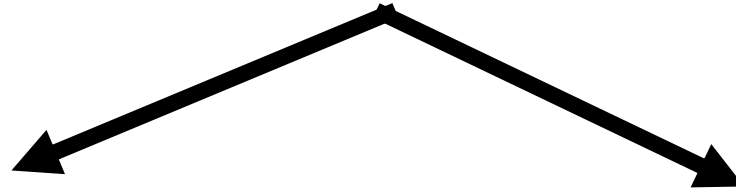
Ce sont les différentes valeurs possibles d'une variable.

Types de Variable:

- Qualitatives ordinale
- Qualitatives nominale
- Quantitative discrète.
- Quantitative continue.

VARIABLES QUANTITATIVES

Une variable statistique est quantitative si ses modalités sont des nombres



Variable quantitative

discrète:

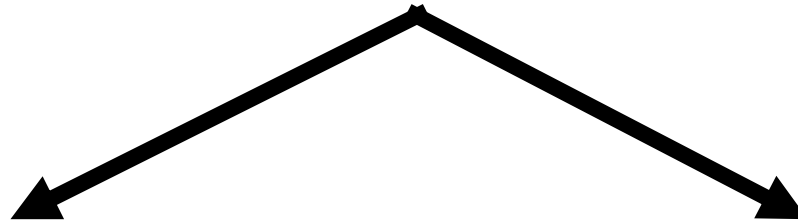
Une variable quantitative est discrète si elle ne peut prendre que des valeurs isolées, généralement entières.

Variable quantitative

continue: Une variable quantitative est continue si ses valeurs peuvent être n'importe lesquelles d'un intervalle réel.

VARIABLES QUALITATIVES

Une variable statistique est qualitative si ses valeurs, ou **modalités**, sont des codes



Variable qualitative nominale :

C'est une variable qualitative dont les modalités ne sont pas ordonnées.

Variable qualitative ordinale

C'est une variable qualitative dont les modalités sont naturellement ordonnées

- **TAILLE ou EFFECTIF TOTAL**
Nombre d'individus de la population ou de l'échantillon
- **EFFECTIF D'UNE MODALITE**
Nombre d'individus qui présentent cette modalité particulière
- **FREQUENCE D'UNE MODALITE (souvent notée f_i)**
Effectif de cette modalité divisé par l'effectif total. C'est la proportion d'individus qui présentent cette modalité.

- **Classe** : c'est l'intervalle choisi pour organiser la série à travers un regroupement;
sa largeur est appelée **amplitude**, $(a_{i+1} - a_i)$
son milieu est le **centre de classe** $.(a_{i+1} + a_i)/2$
- **Effectif Cumulé croissant E.C.C**
L'effectif cumulé croissant de X en a est la somme des effectifs dont la variable associée est inférieure à a .
- **Série statistique** :
C'est l'ensemble des données relevées dans une étude ; $(A_i; n_i)$
ou $(x_i; n_i)$ ou $([a_i, a_{i+1}[; n_i)$

SERIE STATISTIQUE

Modalité		Effectif	E.C.C	E.C.D
a1	b1	n1	$C1=n1$	$D1=N$
a2	b2	n2	$C2=C1+n2$	$D2=D1-n1$
a3	b3	n3	$C3=C2+n3$	$D3=D2-n2$
a4	b4	n4	$C4=C3+n4$	$D4=D3-n3$
.....
.....
ak	bk	nk	$ck=c(k-1)+nk$	$Dk=D_{(k-1)}-n_{(k-1)}$
Total		N		

Une étude statistique peut concerner un seul caractère:

Statistique univariée ou unidimensionnelle

Ou deux caractères

Statistique bivariée

et le plus souvent plusieurs caractères

Statistique multivariée.

- L'effectif corrigé ne se calcule que pour une série quantitative continue.
- Uniquement pour tracer l'histogramme et le calcul du mode (paramètre de position) on doit procéder à la correction des effectifs.
- L'effectif corrigé d'une classe est égal au rapport de l'effectif de la dite classe sur la largeur de la classe.

$$\text{Effectif corrigé} = \frac{\text{Effectif}}{\text{Amplitude}}$$

$$ni^{(c)} = \frac{ni}{a_{i+1} - a_i}$$

VARIABLE QUALITATIVE

Nominale

Ordinale

Effectifs ou Fréquences

Diagramme en
barres

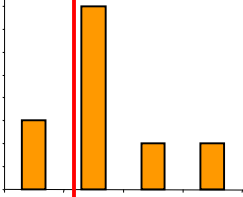


Diagramme en
barres

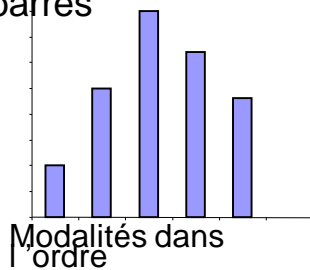


Diagramme
circulaire



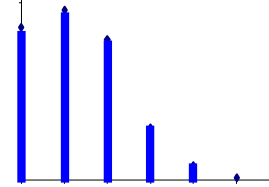
VARIABLE QUANTITATIVE

Discrète

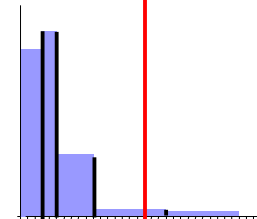
Continue

Effectifs ou Fréquences

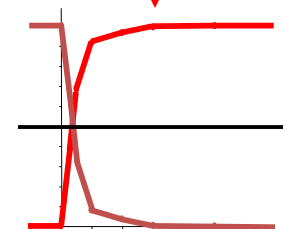
Diagramme en
bâtons



Histogramme



Courbes cumulatives des effectifs ou des fréquences



Un paramètre statistique permet de résumer par une seule quantité numérique une information contenue dans une distribution d'observations.

Ce sont des indicateurs statistiques qui nous renseignent sur la distribution des données :

- La **tendance centrale** ou **position** indique l'ordre de grandeur des données.
- La **dispersion** mesure la variation des données autour de la valeur centrale

- **La Moyenne arithmétique** . (Moyenne harmonique , moyenne géométrique)
- **La médiane**: Valeur qui partage la série ordonnée en deux groupes de même effectif (50% des observations lui sont \leq)
- **Les quartiles** Q1, Q2, Q3 partagent la série ordonnée en 4 groupes de même effectif
- **Le mode**: C'est la valeur la plus fréquente de la série

- L'étendue
- L'intervalle interquartile
- L'écart absolu moyen
- L'écart absolu médian
- La variance
- L'écart type
- Le coefficient de variation