

INTRODUCTION

Objectif : recueillir et analyser des données

Application à tous les secteurs :

- **Industriels** : fiabilité des matériels, contrôle qualité, analyse d'expériences, planification, prévision, ...
- **Sciences humaines** : sondage, enquêtes d'opinion, ...
- **Economie finance** : étude de marché, ...
- **Santé** : épidémiologie, pharmacologie, ...

Le travail du statisticien :

1. Conception
2. Collecte
3. Apurement
4. **Traitement**

INTRODUCTION

Les statistiques : Ensemble de résultats obtenus à la suite de nombreuses observations d'un même phénomène

La statistique : Méthodes scientifiques pour traiter et interpréter les statistiques

Exemple : Relevés quotidiens sur 5 ans du débit d'une rivière.
Objectifs : faire apparaître une périodicité calculer un débit moyen, prévoir la valeur max de la crue annuelle ...

Les probabilités : Outils mathématiques de la statistique

Différence entre statistique et probabilités : Quelle est la probabilité d'obtenir 5 piles sur 6 lancers d'une pièce?

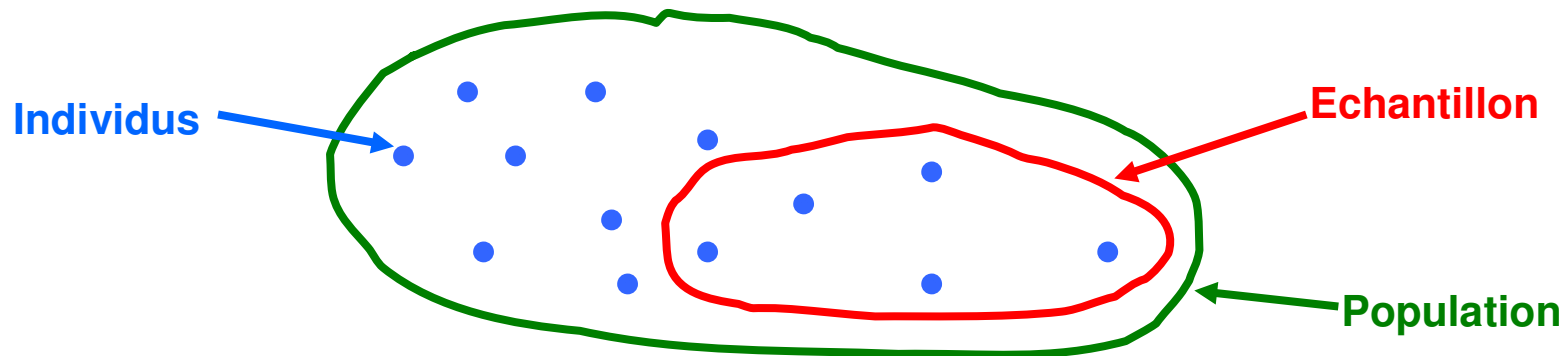
Proba : $C_6^5 (p)^5(1-p)^{6-5}$ (valeur exacte)

Stat : on effectue plusieurs fois une série de 6 lancers et on calcule la moyenne (valeur approchée)

INTRODUCTION

Population et échantillon

- **Individus** : objets équivalents sur lesquels on observe les mêmes caractéristiques
- **Population** : Ensemble des individus
- **Echantillon** : Sous-ensemble de la population



- **Recensement** : Etude de tous les individus d'une population
- **Sondage** : Etude d'une partie de la population

INTRODUCTION

Les variables

Chaque individu est décrit par un ensemble de caractéristiques appelées **variables**

Les variables sont classées suivant leur nature :



INTRODUCTION

Tableau de données



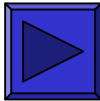
		Variables					
		1	2	...	j	...	p
Individus	1						
	2						
	...						
	i				x_i^j		
	...						
	n						

Variable 1	1	2	...	i		
Effectifs	n_{11}			n_{1i}		
Variable 2	1	2	j	
Effectifs	n_{21}				n_{2j}	
...						
Variable p	1	2	k
Effectifs	n_{p1}					n_{pk}

INTRODUCTION

La statistique descriptive

Objectif : Phase exploratoire pour synthétiser, résumer, structurer l'information contenue dans les données

- Visualisation  
- Résumé numérique 
- Analyse de données
 - Méthodes de classification : réduire la taille de l'ensemble des individus en formant des groupes homogènes
 - Méthodes factorielles : réduire le nombre de variables en les résumant par un petit nombre de composantes synthétiques

INTRODUCTION

La statistique inférentielle

Objectif : Etendre des propriétés constatées sur un échantillon à l'ensemble de la population et valider ou infirmer des hypothèses a priori ou formulées à l'issue de la phase exploratoire

Exemples :

- Estimation d'un paramètre
- Vérification d'une hypothèse
- La modélisation et la prévision statistique

Importance capitale des notions de convergence des variables aléatoires

INTRODUCTION

Exemple de l'estimation d'une moyenne

Une même grandeur m est mesurée n fois de suite par un observateur. Le contexte expérimental rend fluctuantes ces mesures et on obtient n valeurs différentes x_1, \dots, x_n pour cette grandeur.

Quelle est la vraie valeur m ?

On admet que ces valeurs constituent des réalisations indépendantes d'une même variable aléatoire d'espérance m . La loi des grands nombres montre que la moyenne de l'échantillon,

$$\bar{X} = \frac{x_1 + \dots + x_n}{n}$$

constitue une bonne approximation de m . On dit que la moyenne est une estimation de m .

Estimation liée à l'échantillon

On peut donc considérer que \bar{X} est la réalisation d'une variable aléatoire \bar{X} . Les probabilités fournissent un intervalle de confiance dans lequel il y a par exemple 95% de chance de trouver la vraie valeur m .

INTRODUCTION

Exemple d'un test d'hypothèse en contrôle qualité

Un client commande à son fournisseur des lots de pièces dont la qualité est spécifiée dans le cahier des charges à 4%. Avant de livrer, le fournisseur effectue un contrôle sur un échantillon de 50 pièces et trouve 3 pièces défectueuses, *i.e* 6% de son échantillon.

Doit-il livrer quand même?

Le raisonnement est le suivant. Le nombre théorique de pièces défectueuses parmi 50 suit une loi binomiale $B(50,0.04)$ donc la probabilité de trouver 3 pièces défectueuses est $C_{50}^3 0.04^3 0.96^{47} = 0.32$, *i.e* 32%. Cette probabilité étant assez élevée aux yeux du fournisseur (il avait une chance sur trois de trouver 3 pièces défectueuses dans l'échantillon), il décide de livrer. Le client acceptera-t'il?

INTRODUCTION

Modélisation et Prévision

La modélisation consiste généralement à chercher une relation approximative entre plusieurs variables, cette relation étant le plus souvent linéaire,

$$Y = aX + b$$

où a et b sont des constantes à estimer.

Dans le cas de variables aléatoires quantitatives, on parle de **régression linéaire**. Le modèle obtenu sert ensuite à faire de la **prévision** en des points non observés.

