

Quel plan experimental mettre en oeuvre?

1.2. Les plans à groupes (échantillons) appariés (= non indépendants)

2 échantillons A et B sont appariés si chaque observation de A est liée à une observation de B. Chaque couple de valeurs forme alors une paire : A_1-B_1 , A_2-B_2 , A_3-B_3 ...
 → Les échantillons appariés sont souvent de la même taille.

Exemples de situations fréquentes:

- n_1 et n_2 individus différents mais liés entre eux (ex: jumeaux 1 et 2, frère-sœur)

Ex: n couples d'individus homogènes. Le 1er individu du couple = trait. A, 2^e = contrôle.

- 1 seul échantillon de n individus exposés successivement à toutes les modalités de la VI. La VD est mesurée plusieurs fois sur les mêmes individus → « mesures répétées ».

Ex: chacun des n individus est mesuré avant puis après avoir reçu le traitement A (chaque individu est son propre contrôle).

On s'intéresse à la moyenne des différences entre paires de valeurs

Indiv	Contrôle	Trait. A
1	y1.1	y1.2
2	y2.1	y2.2
...
45	y45.1	y45.2

Quel plan experimental mettre en oeuvre?

2) Les plans à plusieurs variables indépendantes : plans factoriels.

Permet de tester toutes les combinaisons possibles de modalités (≥ 2) de différentes VI (≥ 2).

1 combinaison = 1 condition expérimentale à laquelle es soumis un certain effectif d'indiv.

ex: un plan factoriel de 3 facteurs à 2 modalités $\rightarrow 2^3 : 8$ conditions expérimentales

Avantages :

- Tester les effets Principaux de chaque VI sur la VD
- Tester les effets d'Interactions entre plusieurs VI sur la VD (ex: action synergique)
- Nécessite moins d'individus qu'un étude réalisée avec chacune des modalités séparée

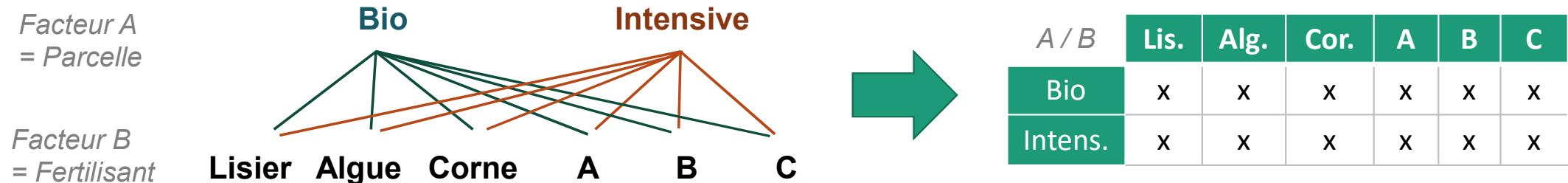
→ ANOVA / modèles linéaires : Décomposition de la variation de la VD; part imputable au effets principaux (et d'interaction) de chaque VI + part due au hasard.

NB: Le plan peut être à groupes indépendants, à groupes appariés, ou encore mixte.

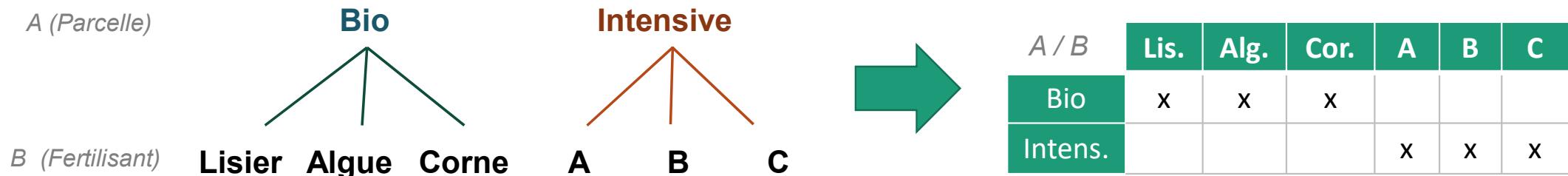
Quel plan experimental mettre en oeuvre?

Relation entre Facteurs : croisée ou emboîtée (= hiérarchisée) ?

Plan croisé : chaque modalité du 1^{er} facteur est croisée avec chaque modalités du 2nd
 → Toutes les combinaisons sont testées : $A_2 * B_6 = 12$ conditions expérimentales



Plan emboîté : Un facteur B est emboîté dans un facteur A lorsque chaque modalité de B n'est associée qu'à une seule modalité de A (facteur emboitant) → On note $B_6 < A_2 >$.



Quel plan experimental mettre en oeuvre?

Plan complet ou incomplet?

Ex: *VD = qualité épis et 2 facteurs croisés: variété (traditionnelle vs OGM) * type d'engrais (A vs B)*

.

Plan complet

	A	B
Trad.	X	X
OGM	X	X

Toutes les combinaisons possibles sont testées (ici: 2*2)

Plan incomplet

	A	B
Trad.		X
OGM	X	X

Certaines combinaisons ne sont pas testées :

- **Par accident** (ex: maladie tuant variété Trad. en A)
- **Par choix** → **Plans fractionnaires** : réduire coûts en se concentrant sur une fraction des conditions expérimentales (effets les plus intéressants).

Quel plan experimental mettre en oeuvre?

Plan équilibré ou déséquilibré?

Ex: *VD = qualité épis et 2 facteurs croisés: variété (traditionnelle vs OGM) * type d'engrais (A vs B)*

→ Effectifs = nombre d'individus statistiques/unités expérimentales (ex: pieds de maïs)

Plan équilibré

	A	B
Trad.	60	60
OGM	60	60

Plan déséquilibré

	A	B
Trad.	0	256
OGM	302	480

Chaque condition expérimentale compte le même nombre d'individus / d'observations.

Les effectifs sont inégaux.

NB : traitement stat. des plans équilibrés généralement plus facile (nombreuses méthodes) et efficace (plus puissantes) que pour les plans déséquilibrés...

Quel plan experimental mettre en oeuvre?

Autres types de plans : Type 1, 2 et mixte ; qu'est ce que c'est ?

Dépend de la connaissance du **type de facteur(s) principal(aux)** :

- **Plan de type I** : Uniquement des facteurs principaux à effets fixes (systématiques)
→ ANOVA « modèle I » (effets portent sur la moyenne de la VD)
- **Plan de type II** : Uniquement des facteurs principaux à effets aléatoires
→ ANOVA « modèle II » (effets portent sur la variance de la VD)
- **Plan mixte** : mélange des 2 plans précédents (effets fixes ET aléatoires)
→ ANOVA « modèle III »

... Pour en savoir plus ...



Plan en étoile

Carré latin

Randomisation totale

Cross over

Split splot

Bloc

Mais aussi :

Petit abrégé pour mieux comprendre la notion de méthode expérimentale et ses enjeux méthodologiques (PDF).

https://www.researchgate.net/publication/44164714_Petit_abrege_pour_mieux_comprendre_la_notion_de_methode_experimentale_et_ses_enjeux_methodologiques

Plan du cours

Partie 1. Démarche scientifique et introduction aux statistiques

1. Introduction : De la problématique à l'hypothèse de recherche
2. **Méthodologie de recherche : Protocole et planification**
 - Les variables
 - Population, échantillon, individu statistique
 - Type d'investigation : Etude expérimentale vs observationnelle
 - Introduction aux plans expérimentaux
 - **Introduction aux stratégies d'échantillonnage**
3. Vers l'analyse statistique des données



Plan d'échantillonnage

45

Population :



Echantillonnage
(9 individus)

Représentativité : Quel échantillon choisir ?



quasi que
des femmes



quasi que
des rouges



quasi que des enfants



OK

Plan d'échantillonnage

1. L'échantillonnage non probabiliste :

- Les **n** individus ne sont pas sélectionnés aléatoirement (comporte un part d'arbitraire)
- On ne connaît pas la probabilité pour 1 individu quelconque d'être inclus dans l'échantillon, certains peuvent n'avoir aucune chance d'être sélectionné.

Conséquence: pas possible d'évaluer la représentativité des tels échantillons, ni la précision des estimations qui seront réalisées à partir de ces derniers (ex: moyennes, %).
→ Résultats faiblement généralisables (faible validité externe).

« **Avantage** » : simple et peu couteux à réaliser..

Type de plan d'échantillonnage principalement utilisé en Sciences sociales et Santé.

Plan d'échantillonnage

Échantillon de commodité/convenance :



Les individus sont choisis pour des raisons pratiques de disponibilité, accessibilité et coût.

Ex : Un enquêteur interroge les 20 premiers passants de la rue commerçante chic où il se trouve sur l'intérêt des instituts de beauté.

Biais probables : De sélection (spatiotemporel) et de représentativité.

En quoi les passants de cette rue à ce moment de la journée sont représentatifs de la pop. ?

Plan d'échantillonnage

Échantillon de volontaires :

Consiste à faire appel à des volontaires pour qu'ils fassent partie de l'échantillon.

Pour des résultats valides, l'échantillon doit être très semblable à la population.

Exemple 1 : Un fabricant de yaourt demande à ses consommateurs de remplir un questionnaire de satisfaction en ligne.



Biais probables : Produit généralement un échantillon d'individus polarisés.

Volontaires = souvent les personnes ayant les opinions les plus extrêmes.

→ peu représentatif de la population (majorité silencieuse non représentée).



Plan d'échantillonnage

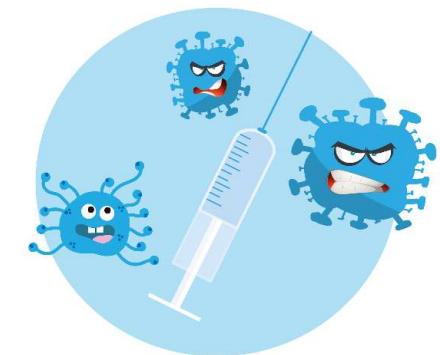
Échantillon de volontaires :

Consiste à faire appel à des volontaires pour qu'ils fassent partie de l'échantillon.

Pour des résultats valides, l'échantillon doit être très semblable à la population.

Exemple 2 :

Une industrie pharmaceutique lance un appel à participation pour tester l'efficacité d'un vaccin (essai clinique).



Biais probables : Généralement négligeables car ici l'étude ne porte pas sur une opinion mais sur la réaction du corps à un traitement.

→ L'échantillon est représentatif de l'ensemble des «corps» des individus de la population.

Plan d'échantillonnage

2. L'échantillonnage probabiliste (aléatoire) :

- Les n individus sont sélectionnés aléatoirement / indépendamment les 1 des autres
→ Pas d'intervention du chercheur; seul le hasard régit l'inclusion ou non d'individus.
- Les individus ont tous la même probabilité de faire partie de l'échantillon, cette probabilité est non nulle et connue

Conséquence:

- L'échantillon est représentatif de la population source
- Généralisation possible des résultats à la pop (statistiques inférentielles)
- On peut évaluer la précision de ces estimations

Plan d'échantillonnage

... Il existe différentes méthodes d'échantillonnage probabiliste ...

On se concentre ici sur **les plus utilisées en (agro)écologie** afin de **sélectionner des sites d'échantillonnage** pour réaliser des études d'expérimentation / observation *in situ*.

Ici: **1 site** = un individu statistique (point, quadra, transect...)

1 échantillon = un ensemble de **n** site

→ Cet échantillon doit être représentatif de tous les terrains possibles où l'expérience / observation aurait pu être menée si les moyens l'avaient permis (**population de sites**) afin de permettre une généralisation des résultats obtenus.

→ Pour être représentatif, il doit être aléatoire.

Plan d'échantillonnage

Échantillonnage aléatoire simple :

Sélection au hasard et de façon indépendante de la position des sites dans une fenêtre spatiale → Chaque point de l'espace étudié a une probabilité égale d'être sélectionné.

Procédure :

1. Bien définir sa population d'intérêt (*qqsoit le plans choisi*)
2. Déterminer la taille n de l'échantillon
3. Tirage aléatoire* des coordonnées des n sites (ex: avec R) → échantillon



2x2 km

* (*avec remise* : Permet de maintenir constante la probabilité de sélection de chaque individu statistique/site → important pour les petites populations)

Plan d'échantillonnage

Échantillonnage aléatoire simple :

Avantages :

- Se prête bien aux analyses statistiques...
- Stratégie très efficace (écologie spatiale) permettant de répondre à de nombreuses questions avec même échantillon
- Nécessite pas de connaissance préalable sur les variables

Inconvénients :

- Difficile à disposer dans l'espace et lourd à déployer (distances entre sites)
- Certains sites peuvent tomber dans des zones hors contexte d'étude (ex: routes vs habitats naturels, milieu aquatique vs terrestre) ou difficile d'accès
- Quand les milieux/habitats sont de surfaces inégales, on aura peu d'échantillons du type rare (échantillonnage insuffisant) et beaucoup de type fréquents (sur-échantillonnage)



Plan d'échantillonnage

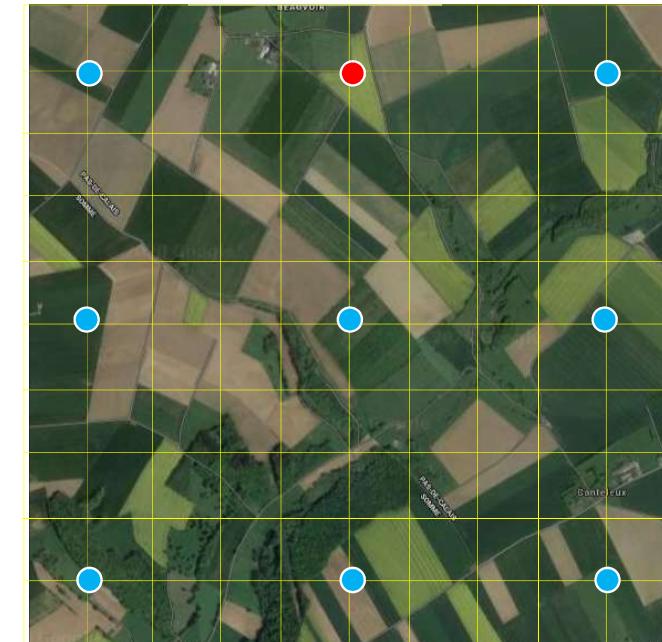


Échantillonnage systématique :

Les sites sont répartis à intervalles réguliers d'espace et/ou de temps dans la zone d'étude (= équidistants), organisé dans un quadrillage (grille 2D) ou le long d'un transect (1D).

Procédure :

1. Même étapes 1 et 2 que EAS
2. Déterminer le pas/intervalle d'échantillonnage
(écartement *entre chaque site*)
3. Fixer l'orientation du transect ou de la grille (aléatoire?)
4. Tirer au hasard les coordonnées d'un 1^{er} site (EAS)
5. Positionner les n-1 sites suivants selon l'intervalle déterminé en 3



(ex: 4 intersections = 800 m)

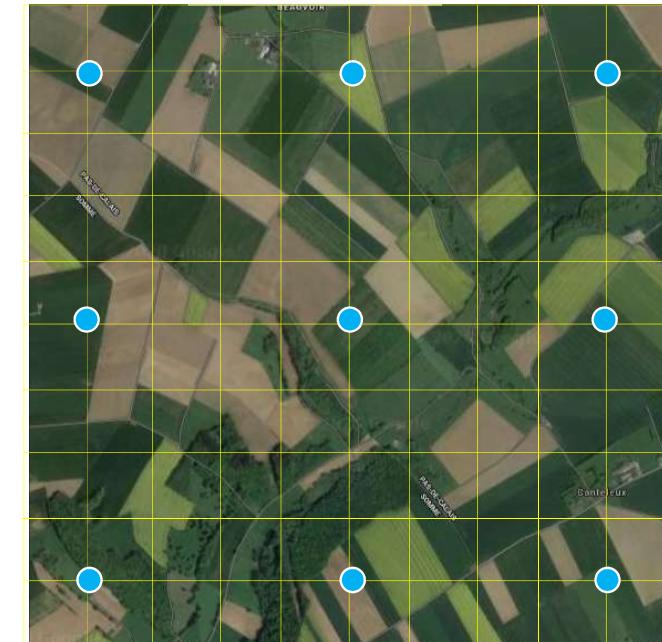
Plan d'échantillonnage



Échantillonnage systématique :

Avantages :

- *Idem que aléatoire simple*
- + Facilité de mise en œuvre et de planification
- + Permet fournir cartes d'occupation de l'espace (temps) d'espèces végétales et/ou animales
- + Les pièges pour espèces difficilement observables (invertébrés, mammifères) sont souvent positionnés de manière systématique sur 1 secteur ou le long de gradients



(ex: 4 intersections = 800 m)

Inconvénients :

- *Idem que aléatoire simple*

Plan d'échantillonnage

Échantillonnage aléatoire stratifié :

consiste à subdiviser une population hétérogène en sous-pops (strates) plus homogènes, mutuellement exclusives et collectivement exhaustives*, selon un critère de stratification (1 ou plusieurs facteurs principaux).

* (chaque individu appartient à 1 strate et 1 seule!)

Procédure :

1. Même étapes 1 et 2 que EAS

2. Découpage en strates selon les modalités d'1 ou + facteur(s)

3. Choix nombre d'individus par strate (proportionnalité?)

4. Tirage aléatoire coord. de n_i sites dans la strate $i \rightarrow \sum n_i = n$ taille totale échantillon

Echantillon n = 9 (proportionnelle)

Stratification selon le type de culture :

 - A
 - B
 - C



Echantillon n = 9 (proportionnel)

Stratification selon le type de culture :

- A=29%, $n_A=3$
- B=57%, $n_B=4$
- C=14%, $n_C=2$

→ $\Sigma n_i = 9$

Plan d'échantillonnage

Échantillonnage aléatoire stratifié :

Avantages :

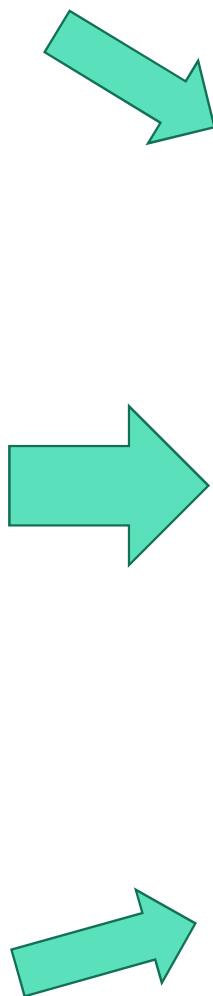
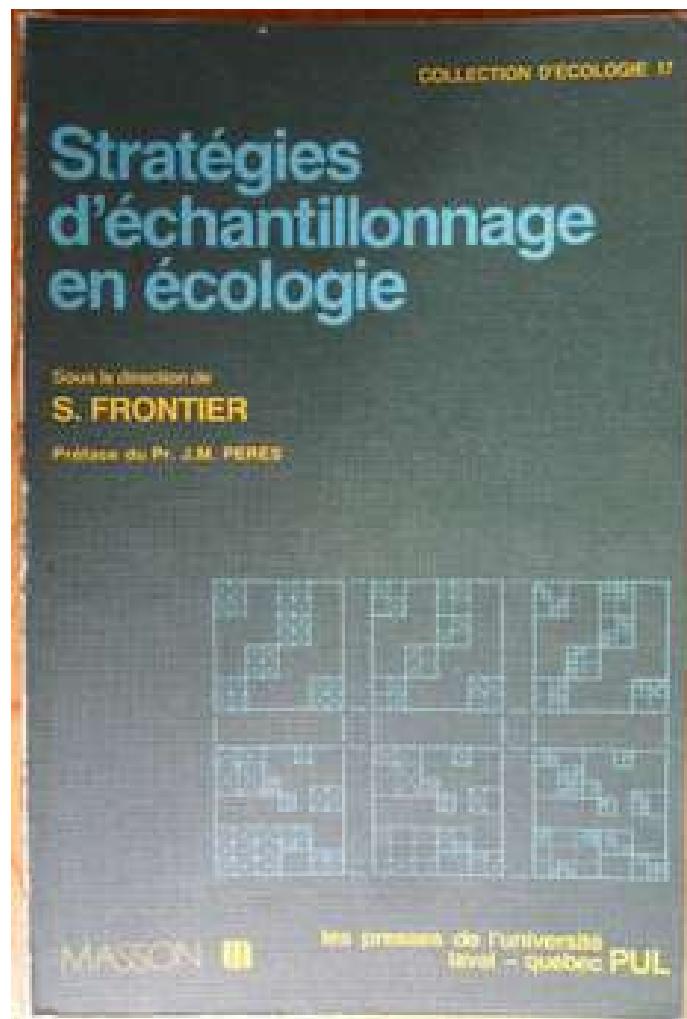
- Stratégie très efficace et économique quand l'objectif est précis
- Échantillonnage probabiliste : traitements statistiques possibles
- Adapté à l'étude de l'effet d'un ou plusieurs facteurs principaux car permet de maîtriser les facteurs secondaires si nécessaire



Inconvénients :

- Nécessite bonne connaissance préalable sur la/les variables sujettes à l'échantillonnage
- Important travail de préparation de l'échantillonnage
- Devient vite complexe/coûteux si cherche à répondre à plusieurs questions en même temps

... Pour en savoir plus...





Plan d'échantillonnage

Chaque étudiant d'une université possède un n° d'identification. Une société de sondage sélectionne au hasard un identifiant, puis successivement tous ceux ayant un intervalle d'identification de 50. Tous ces étudiants sont interrogés sur leur pratiques numériques.

De quelle méthode d'échantillonnage s'agit il?

- A. De commodité
- B. Aléatoire simple
- C. Aléatoire stratifié
- D. Systématique
- E. Volontaire

✓ D

Plan d'échantillonnage



Pour évaluer le niveau de satisfaction de ses clients, un restaurant laisse à disposition sur les tables un bref questionnaire. Les clients qui le souhaitent peuvent le remplir.

De quelle méthode d'échantillonnage s'agit-il?

- A. De commodité
- B. Aléatoire simple
- C. Aléatoire stratifié
- D. Systématique
- E. Volontaire

✓ E

Plan d'échantillonnage



Une grande entreprise mène une enquête auprès de 100 de ses salariés en prélevant de façon aléatoire 10 gestionnaires et 90 employés.

De quelle méthode d'échantillonnage s'agit-il?

- A. De commodité
- B. Aléatoire simple
- C. Aléatoire stratifié
- D. Systématique
- E. Volontaire

✓ C

Plan d'échantillonnage



Pour sélectionner 5 de ses employés, le directeur de cette entreprise associe à chacun d'eux un nombre écrit sur une balle de ping-pong. Il met les balles dans un sac et, sans regarder, tire au hasard 5 boules. Les employés associés à ces nombres constituent l'échantillon.

De quelle méthode d'échantillonnage s'agit-il?

- A. De commodité
- B. Aléatoire simple
- C. Aléatoire stratifié
- D. Systématique
- E. Volontaire

✓ B

Plan du cours

63

Partie 1. Démarche scientifique et introduction aux statistiques

1. Introduction : De la problématique à l'hypothèse de recherche
2. Méthodologie de recherche : Protocole et planification
 - Les variables
 - Population, échantillon, individu statistique
 - Type d'investigation : Etude expérimentale vs observationnelle
 - Introduction aux plans expérimentaux
 - Introduction aux stratégies d'échantillonnage
3. Vers l'analyse statistique des données

Quelles analyses statistiques?

Statistique : ensemble de méthodes pour **collecter, organiser, présenter, résumer, et analyser** des données, pour en **tirer des conclusions** et de prendre des décisions.

1. Descriptive : **Synthétiser, résumer et structurer** l'information contenue dans les données (échantillon), mettre en évidence des propriétés et suggérer des hypothèses (*tableaux, graphiques, distribution, indicateurs numériques, relations*).

Rôle mineur des probabilités.



1.bis. Exploratoire : **Visualiser et explorer** jeux de données complexes, identifier relations entre VI (*graphiques multidimensionnels, ex: ACP, AFC*).

2. Inférentielle :

- **Généraliser** aux populations les conclusions tirées à partir d'observations sur échantillons (*estimations de paramètres, décisions issues de tests d'hypothèses, relations*)
- **Expliquer et prédire** des données quantitatives ou qualitatives à partir de modèles (*regression, classification*).

Nécessite de définir des modèles probabilistes et savoir gérer les risques d'erreurs (Partie 2)

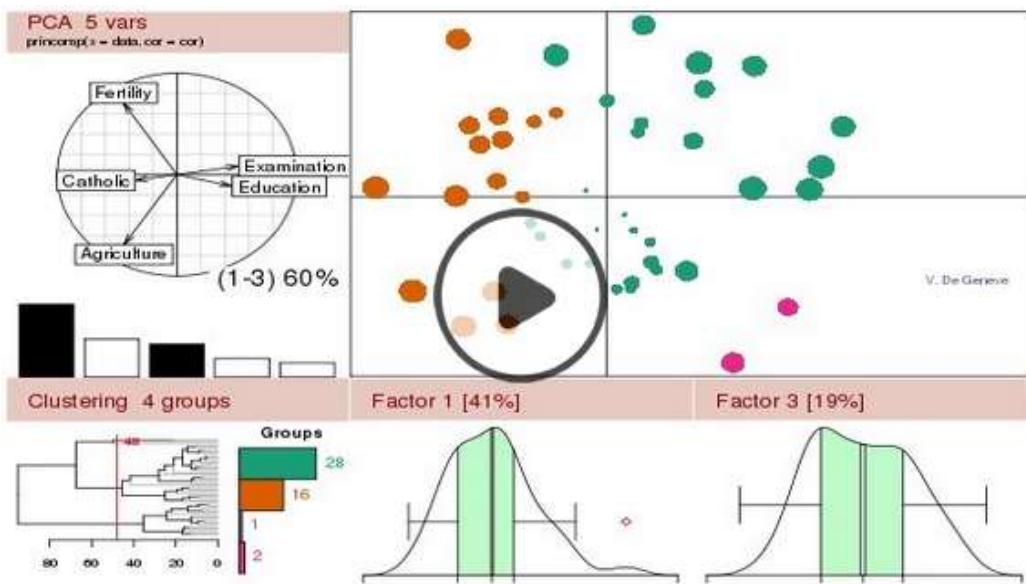
Plateforme Fun Mooc

(cours gratuits en ligne)

<https://www.fun-mooc.fr>

Introduction aux statistiques avec R

... inscrivez-vous !!! ...



université PARIS-SACLAY

[**f**](#) [**in**](#) [**✉**](#)

Inscription
Du 21 juillet 2022 au 21 octobre 2022

Cours
Du 12 septembre 2022 au 26 octobre 2022

Langues
Français

S'inscrire maintenant

FIN PARTIE 1