

TD 1 : Analyse des données

ANOVA

On veut connaître l'effet de trois types de fertilisants sur la croissance des arbres d'une plantation.

1. Principe de l'expérimentation

Extraire 3 échantillons (groupes) d'arbres et appliquer chaque **fertilisant** pour chaque échantillon : comparer ensuite les moyennes de **croissance annuelle** des arbres.

Variable d'intérêt (variable dépendante)
en cm/an par exemple.
Autres exemples

- Rendement d'un paquet d'action.
- Taux de virus dans le sang.

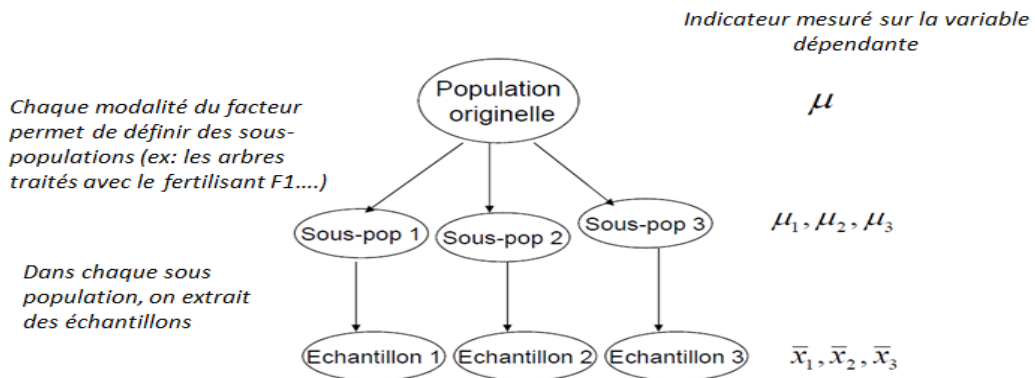
Facteur (variable indépendante)
Type de fertilisant.
Autres exemples

- Stratégie de placement.
- Traitement médical.

L'ANOVA s'applique dès que :

- On veut monter une expérimentation ;
- On veut vérifier l'effet de variables qualitatives sur une variable quantitative.

2. Principe statistique



La problématique de l'ANOVA consiste à utiliser les moyennes observées sur les échantillons pour conclure à des différences significatives sur les moyennes (espérance mathématique) dans les sous-populations.

Ex : Pour la plantation, tous les fertilisants sont-ils équivalents, ou bien y a-t-il un qui soit meilleur (moins bon) que les autres... ?

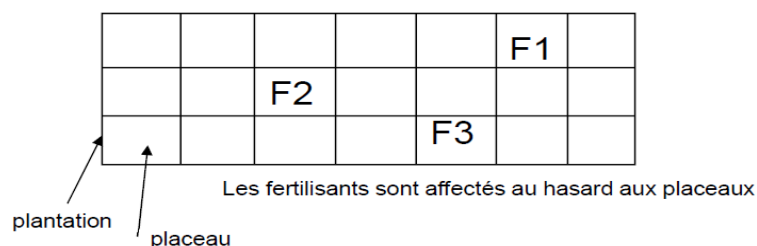
3. Problèmes pratiques

A. Plan d'expérimentation

Dans la pratique, plusieurs problèmes peuvent corrompre les résultats...

Ex: On ne peut pas donner un fertilisant à un arbre, il y a des problèmes de diffusion aux arbres voisins (pluie, vent...)

- ⇒ Une solution possible serait d'effectuer un maillage de la plantation (on définit ainsi des placeaux), on applique un fertilisant aux arbres qui sont dans le même placeau.



Quelques définitions

- ☞ individu statistique : un arbre de la plantation
- ☞ population : les arbres de la plantation
- ☞ échantillon expérimental : les arbres dans les 3 placeaux
- ☞ unité expérimentale : un placeau

B. Les facteurs non-contrôlés

Si la plantation est grande, différents facteurs peuvent perturber l'expérimentation :

- Différences climatiques, il peut y avoir des zones de micro-climat ;
- Facteurs génétiques : les graines n'ont peut-être pas tous la même provenance et des arbres voisins (sur le même placeau) ont toutes les chances de provenir du même lot ;
- Le sol n'est pas de même qualité dans toute la plantation.

Facteurs non-contrôlés, on sait qu'ils peuvent perturber les résultats mais on ne sait pas les prendre en compte explicitement dans l'analyse.

Le rôle du plan d'expériences est de définir au mieux l'expérimentation (ex: répartir les traitements sur les placeaux) de manière à annihiler le rôle des facteurs non-contrôlés.

C. Les facteurs de blocs

Certains facteurs (que l'on connaît cette fois-ci) peuvent perturber les résultats, il appartient au statisticien de les maîtriser au mieux.

Ex: Pour évaluer un médicament, les facteurs de blocs peuvent être l'âge (jeune, adulte, vieux), le sexe (homme, femme)... si on doit évaluer 2 médicaments, l'un est administré uniquement aux hommes, l'autre aux femmes, la validité des résultats devient sujette à caution...

Même si ce n'est pas l'objectif (évaluer l'efficacité des médicaments selon le sexe), il est important d'en tenir compte pour évaluer les résultats (qui sont de mesurer l'efficacité du médicament sans distinction de sexe)

Il est également possible de supprimer l'effet de bloc par un plan d'expériences approprié, par ex. en mettant le même nombre d'hommes et de femmes dans chaque échantillon

D. Facteurs fixes et facteurs aléatoires

Dans l'exemple des fertilisants, le nombre de modalités du facteur est faible (3 fertilisants qui sont tous connus). Que faire quand il est innombrable, ou que l'on ne veut en traiter que quelques-uns pour inférer sur les autres.

E. Récapitulatif sur les facteurs de variation

