

# Biométrie

Mm. Mebrek N,

# Pourquoi échantillonner ?

- Impossibilité d'accéder
  - À tous les individus d'une population
  - À la totalité d'une aire
- => On procède donc par inférence
  - Echantillonnage
  - Sondage

# Deux grandes stratégies

- Aléatoire
- Simple
- Stratifié
- En grappe ou par degrés
- Systématique

# Echantillonnage aléatoire simple

- Définition

- ❖ Les individus de la population sont tous équivalents
- ❖ Le nombre d'individus à échantillonner est déterminé à l'avance
- ❖ Chaque individu de la population a la même probabilité  $a$  priori d'être choisi
- ❖ Le choix d'un individu ne favorise ni ne défavorise le choix ultérieur d'aucun autre individu de la population (tirages indépendants)

# Les individus ou unités d'échantillonnage

- Naturels
  - Animaux, plantes individualisées
- Arbitraires
  - Unités de surface, de volume, de poids
    - 0.25 m<sup>2</sup> de prairie
    - 1dm<sup>3</sup> d'eau dans un étang
    - 1k de sol
- Attention alors !
  - Population biologique
  - Population statistique

# Deux mode de tirage

- Avec remise
  - Ou non exhaustif
  - La probabilité de sélection reste constante au cours de l'échantillonnage
- Sans remise
  - Ou exhaustif
  - La probabilité de sélection s'accroît au cours de l'échantillonnage

# Une approximation

- Dans les très grandes populations, on considère souvent l'échantillonnage comme avec remise, même lorsqu'il n'y a pas remise
- Dans les populations plus petites, il y aura lieu de prendre en compte le **taux de sondage**  $f = n/N$

# Paramètres et estimateurs

- La moyenne est un **estimateur sans biais** de la moyenne de la population

$$E(\bar{x}) = \mu$$

- La variance  $s^2$  est un estimateur **biaisé par défaut** (mais **asymptotiquement** sans biais) de la variance de la population

$$E(s^2) < \sigma^2$$

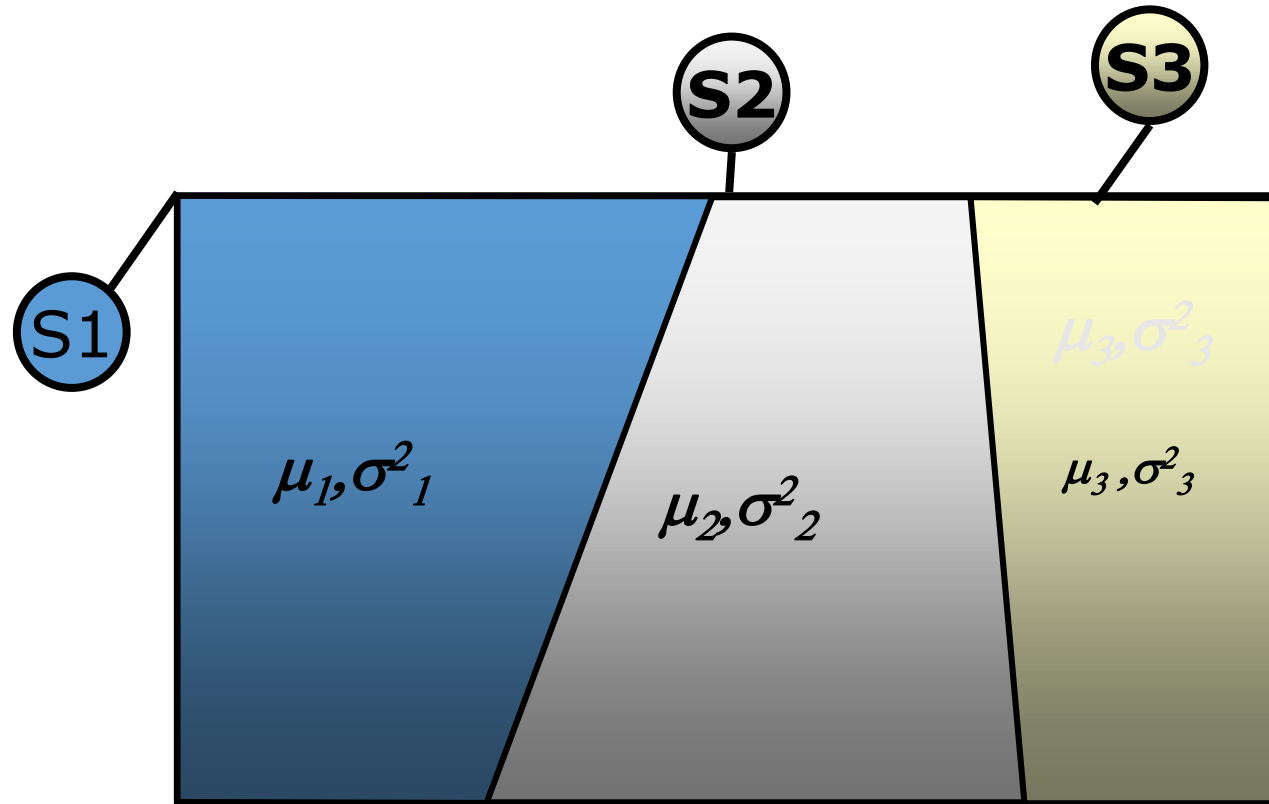


# 3. L'échantillonnage stratifié

Du bon usage des strates

# Une nouvelle vision de la population

- Et des paramètres



# Définition des strates

- Les strates forment une **partition** de la population
- C'est à dire que leurs intersections sont deux à deux vides (elles sont disjointes)
- Leur réunion est la population totale

# Intérêt des strates

- Comment juger de l'intérêt de la stratification ?
- Par analyse de variance
- Une technique qui permet de comparer la variance inter-strate avec la variance intra-strate
- Plus le  $F$  est grand, plus la stratification est intéressante
- A l'inverse, si  $F$  est non significatif, la stratification est dépourvue d'intérêt

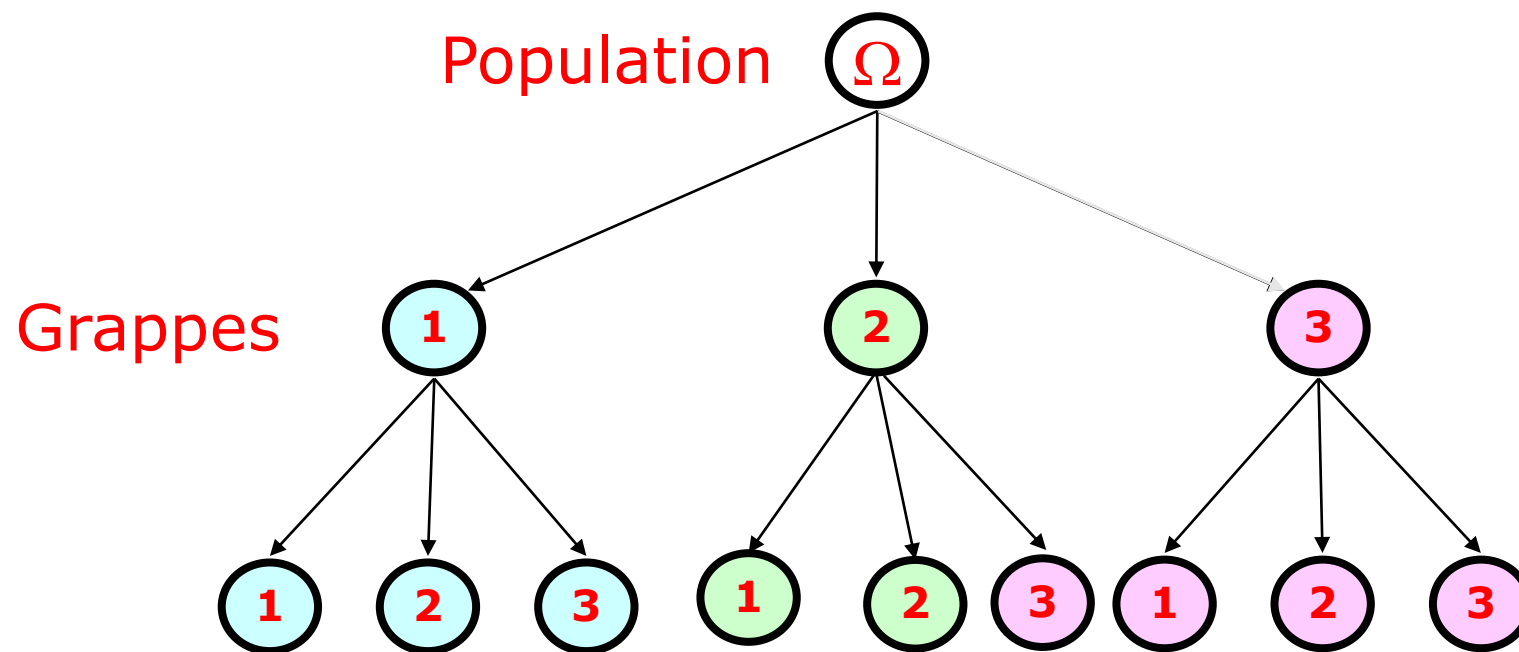
# 4. L'échantillonnage en grappes

Ou échantillonnage du premier degré  
(échantillonnage par degrés)

# Définition

- La population peut être subdivisée en unités primaires ou grappes
- Chaque grappe contient un certain nombre d'individus ou grains
- Le tirage au hasard s'effectue en deux phases
  - Choix de  $m$  grappes
  - Choix de  $n$  grains par grappe
- Analogie : strates très nombreuses, on ne peut les sonder toutes

# Un schéma hiérarchique



# 5. Autres plans d'échantillonnage

Échantillonnage par degrés,  
échantillonnage en différentes occasions,  
échantillonnage par régression



# Echantillonnage par degrés

- Généralisation de l'échantillonnage en grappes
- Echantillonnage en grappe = échantillonnage du premier degré
- Echantillonnage du second degré :
  - On tire au hasard des unités primaires
  - Dans chaque unité primaire on tire au hasard des unités secondaires
  - Dans chaque unité secondaire des unités tertiaires (grains)
- En anglais : *cluster sampling*

# Echantillonnage à différentes occasions

- On tire au hasard un certain nombre d'individus dans une population
- On les repère
- On mesure une caractéristique plusieurs fois (occasions)
- Exemples : croissance sur des animaux ou plantes marquées
- Analyse : « mesures répétées » (repeated measures)

# Echantillonnage par régression

- On mesure une caractéristique peu coûteuse  $x$  sur un très grand nombre  $N$  d'individus
- Sur un sous-échantillon aléatoire de taille  $n$ , on mesure une autre caractéristique, très coûteuse,  $y$
- Ce sous échantillon permet d'estimer le coefficient de corrélation entre les deux caractéristiques
- L'estimation précise de la moyenne de  $x$  permet alors de corriger la moyenne de  $y$

# Que faire si la variance des individus est élevée ?

- L'obtention d'une bonne précision est alors extrêmement coûteuse
- Mais la population est peut-être très hétérogène ?
- On peut alors la diviser en **sous populations** plus homogènes
- On gagne alors beaucoup de précision