

**Travaux Pratique sur données Agricoles**  
– Document réalisé par *Youssouf BANCE*–

**Partie A**

**1) Décrire sommairement la base TD1\_BaseList**

La base *TD1\_BaseList* comporte 3 variables et 8922 individus. Les individus sont les ménages agricoles. Les variables sont constituées par la zone agro-écologique, la région et l'identifiant (ID).

**2) Vérifier l'existence de potentiels doublons**

Au moyen de la commande **duplicates report** on verifie que la base ne comporte pas de doublons

**3) Peut-on conclure qu'il s'agit d'une base de sondage ?**

Une base de sondage doit être absolument exhaustive et ne pas comportement de doublons. Notre base respect la deuxième propriété. Cependant, en ce qui concerne l'exhaustivité de notre base on ne dispose pas suffisamment d'information pour conclure.

**Partie B**

**1) Quel est le taux de sondage correspondant ?**

Le taux de sondage est donnée par :

$$Taux = \frac{n}{N} \quad (1)$$

avec :

$n$  : La taille de l'échantillon

$N$  : La taille total de la population

La taille totale de la population etant de 8922 et la taille de l'échantillon est de 485 on obtient un taux de sondage de 5,84%.

**2) estimer la population agricole et la superficie totale cultivée**

Pour l'estimation de la population agricole et de la superficie totale cultivée nous avons utilisée l'estimateur horvitz thompson du total définie par :

$$\hat{Y}_{HT} = \sum_{i=1}^n \frac{y_i}{\pi_i} \quad (2)$$

avec :

$y_i$  :Superficie cultivée (nombre de personne) dans le ménage agricole  $i$

$n$  : La taille de l'échantillon

L'estimation de la population total est de 127575.4 alors que celle de la vrai valeur est de 127786. Sa variance est de 12306056 avec un coefficient de variation de 2,75%. Quant à l'estimation de la superficie on trouve une valeur de 35076.705ha alors que la vraie valeur est de 35323.359ha. Sa variance est de 18121170 avec un coefficient de variation de 12,14%. Nous pouvons deduire que l'échantillon tiré est un bon échantillon.

Le resumé de l'estimation par *SAS* est le suivant :

Table 1: Resumé du SAS

Paramètres	Population	Superficie cultivée
Total	127575.4	35076.705ha
Variance	12306056	18121170
Coefficient de variation	2,75%	12,14%

### 3) Donner également les résultats par zone agro-écologique

Table 2: Résultats pour les zone agro-écologique

Zone	Population	Superficie cultivée
Bassin Arachidie	38520.96	17310.52ha
Casamance	27906.54	6822.111ha
Niayes	16243.56	3334.253ha
Sénégal oriental	16721.85	3619.389ha
Vallé du Fleuve	28182.48	3990.433ha

Source : Calcul de l'auteur

## Partie C

### 1) La taille de l'échantillon global et par strate

En appliquant le taux de sondage on trouve un échantillon de 893

En appliquant le taux de sondage sur la taille de chaque on trouve on trouve le tableau suivant :

Table 3: Taille échantillon par zone agro-écologique

Zone	Taille	Taille échantillon
Bassin Arachidie	2962	296
Casamance	1463	146
Niayes	1386	139
Sénégal oriental	1206	121
Vallé du Fleuve	1905	191

Source : Calcul de l'auteur

### 2) Procéder au tirage stratifié

La commande **sample** de *STATA* permet de procéder au tirage de 10 ménage agricole dans chaque zone-écologique.

### 3) Estimer les résultats au niveau strate et au niveau global

Le tableau suivant donne le résumé de l'estimation de la population et de la superficie par global et zone-écologique :

Table 4: Taille échantillon par zone agro-écologique

Zone	Population	Superficie cultivée
Bassin Arachidie	43380	19337.5
Casamance	23090	5478.5
Niayes	19960	4558
Sénégal oriental	18140	3748
Vallé du Fleuve	27130	4216.8
Niveau global	131700	37338.8

Source : Calcul de l'auteur

## Partie D

Sondage à Probas Inégales de 3 régions en prenant comme critère de pondération le nombre de producteurs de la base.

La commande **gsample** de *STATA* permet de procéder au tirage de 3 régions à probabilités proportionnelles aux nombres d'individus statistiques.

La méthode adoptée pour la réalisation de ce tirage sous STATA est la suivante :

- ☞ Calculer le nombre de personne dans chaque ménage agricole en additionnant non résident et résident (pondération)
- ☞ En appliquant la commande **duplicates drop région,force** on obtient une base dont les individus sont les régions.
- ☞ On applique la commande **gsample 3,[w=ponderation]** pour avoir 3 région obtenus par probabilité inégale proportionnelle à leur taille(pondération)