

Echantillonnage et  
estimation  
avec le logiciel SAS



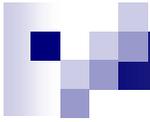
## Plan de la présentation

- Proc SURVEYSELECT et Proc SURVEYMEANS
- Sondage aléatoire simple sans remise
- Plan stratifié
- Plan à probabilités proportionnelles à la taille



## Deux procédures successives

- Tirage d'un échantillon
  - Procédure SURVEYSELECT
  
- Estimation des paramètres
  - Procédure SURVEYMEANS



# Sondage aléatoire simple sans remise



# Tirage d'un échantillon

**PROC SURVEYSELECT**

*DATA = base de sondage*

**METHOD = SRS**

*SAMPSIZE=taille de l'échantillon*

*SAMPRATE=taux de sondage*

*OUT = table échantillon*

**STATS**

**OUTSIZE ;**

*ID liste de variables ;*

**RUN;**



# Estimation des paramètres

## PROC SURVEYMEANS

*DATA = table échantillon*

*{RATE=taux de sondage*

*{TOTAL=taille de la population*

**MISSING**

*statistiques ;*

*WEIGHT variable de pondération ;*

*VAR liste de variables à estimer ;*

*CLASS liste de variables catégorielles à estimer ;*

**RUN;**



# Mots-clés statistiques

MEAN

SUM

VAR

VARSUM

STDERR

STDDEV

CV

CVSUM

CLM

CLSUM

MIN

MAX

RANGE

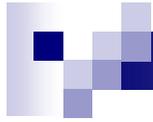
T

DF

SUMWGT

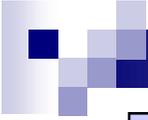
NOBS

NMISS



# Plan stratifié

avec sondage aléatoire simple sans remise  
dans chaque strate



# Allocation proportionnelle

**PROC SORT DATA=** *table base de sondage*;

*BY variables de stratification ;*

**PROC SURVEYSELECT**

*DATA = base de sondage*

*SAMPRATE = f*

*METHOD = SRS*

*OUT = table échantillon*

*OUTSIZE ;*

*STRATA variables de stratification / LIST;*

*ID liste de variables ;*

**RUN;**



# Allocation fixe par strate

## PROC SURVEYSELECT

DATA = *base de sondage*

SAMPSIZE = *n*

METHOD = SRS

OUT = *table échantillon*

OUTSIZE ;

STRATA *variables de stratification / LIST ;*

ID *liste de variables ;*

RUN;

# Allocation différenciée par strate

## PROC SURVEYSELECT

*DATA = base de sondage*

*SAMPSIZE=( $n_1, n_2, \dots, n_p$ )*

*SAMPSIZE=table SAS*

*SAMPRATE=( $f_1, f_2, \dots, f_p$ )*

*SAMPRATE=table SAS*

*METHOD = SRS*

*OUT = table échantillon*

*OUTSIZE ;*

*STRATA variables de stratification / LIST ;*

*ID liste de variables ;*

*RUN;*

## Structure de la table de données secondaires

*Avec SAMPSIZE = table SAS*

*Avec SAMPRATE = table SAS*

Variable de stratification	<i>_NSIZE_</i>
Modalité 1	<i>n1</i>
Modalité 2	<i>n2</i>
...	...

Variable de stratification	<i>_RATE_</i>
Modalité 1	<i>f1</i>
Modalité 2	<i>f2</i>
...	...

# Estimation des paramètres

## PROC SURVEYMEANS

*DATA = table échantillon*

```
RATE =  $f$   
RATE = table SAS  
TOTAL =  $N_h$   
TOTAL = table SAS
```

## MISSING

*statistiques ;*

*STRATA variable(s) de stratification ;*

*WEIGHT variable de pondération ;*

*VAR liste de variables à estimer ;*

*CLASS liste de variables catégorielles à estimer ;*

*BY variable(s) de stratification ;*

RUN;

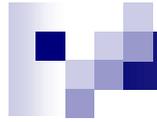
## Structure de la table de données secondaires pour le calcul du facteur $(1-f_h)$

avec **TOTAL** = table SAS

Variable de stratification	<b>_TOTAL_</b>
Modalité 1	$N1$
Modalité 2	$N2$
...	...

Avec **RATE** = table SAS

Variable de stratification	<b>_RATE_</b>
Modalité 1	$f1$
Modalité 2	$f2$
...	...



# Plan à probabilités proportionnelles à la taille



# Tirage d'un échantillon

**PROC SURVEYSELECT**

*DATA = base de sondage*

*METHOD = PPS*

*SAMPSIZE = taille de l'échantillon*

*OUT = table échantillon*

*OUTSIZE ;*

*SIZE variable de taille ;*

*STRATA variable(s) de stratification ;*

*ID liste de variables ;*

**RUN;**



# Estimation des paramètres

## PROC SURVEYMEANS

*DATA = table échantillon*

*RATE = taux de sondage*

*TOTAL = effectif échantillon*

**MISSING**

*statistiques ;*

*WEIGHT variable de pondération ;*

*VAR liste de variables à estimer ;*

*CLASS variable(s) catégorielle(s) à estimer ;*

*STRATA variable(s) de stratification ;*

**RUN;**