

# Formulaire de statistique

**TECHNIQUES D'ECHANTILLONNAGE ET  
D'EXTRAPOLATION**

**TABLE DES MATIERES****SONDAGE ALEATOIRE simple****échantillon****Notation****Définitions****SONDAGE ALEATOIRE simple****extrapolation par le nombre d'individus de la population****calcul de l'extrapolation****SONDAGE ALEATOIRE simple****extrapolation par le ratio****Calcul de l'extrapolation****SONDAGE ALEATOIRE stratifié****extrapolation par le nombre d'individus de la population****Calcul de l'extrapolation****SONDAGE ALEATOIRE stratifié****extrapolation par le ratio séparé par strate****Calcul de l'extrapolation****SONDAGE ALEATOIRE stratifié****extrapolation par le ratio combiné****Calcul de l'extrapolation****SONDAGE ALEATOIRE stratifié****stratification par nuée dynamiques****Algorithme****Sondage aléatoire simple****Tirage systématique ordonné : algorithme d'Ardilly****Sondage aléatoire stratifié****Tirage systématique ordonné : algorithme d'Ardilly****Sondage aléatoire stratifié****allocation de Neyman**

## SONDAGE ALEATOIRE SIMPLE

### ECHANTILLON

#### Notation

$N$	Taille de la population
$y$	variable d'intérêt
$x$	variable auxiliaire
$\bar{x}$	moyenne de $x$ dans l'échantillon
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon
$f$	taux de sondage
$i$	indice d'un individu : $i \in \{1, n\}$

#### Définitions

taux de sondage

$$f = \frac{n}{N}$$

variance corrigée de  $y$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

variance corrigée de  $x$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

covariance de  $x$  et  $y$

$$s_{yx}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

**SONDAGE ALEATOIRE SIMPLE****EXTRAPOLATION PAR LE NOMBRE D'INDIVIDUS DE LA POPULATION**calcul de l'extrapolation

Notation :

$N$	Taille de la population
$T$	total variable d'intérêt dans la population
$\hat{T}$	total estimé de la variable d'intérêt dans la population
$y$	variable d'intérêt
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon
$i$	indice d'un individu : $i \in \{1, n\}$

Formules de calcul :

On calcule la moyenne de la variable d'intérêt dans l'échantillon. C'est le total de la variable d'intérêt dans l'échantillon divisé par le nombre d'individus de l'échantillon.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

le total estimé  $\hat{T}$  de la variable d'intérêt pour la population est égale au produit de la moyenne de la variable d'intérêt dans l'échantillon par l'effectif de la population.

$$\hat{T} = N\bar{y}$$

**SONDAGE ALEATOIRE SIMPLE**

## EXTRAPOLATION PAR LE RATIO

Calcul de l'extrapolation

Notation :

$N$	Taille de la population
$T$	total variable d'intérêt dans la population
$\hat{T}$	total estimé de la variable d'intérêt dans la population
$F$	total variable auxiliaire dans la population
$y$	variable d'intérêt
$x$	variable auxiliaire
$\bar{x}$	moyenne de $x$ dans l'échantillon
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon

Formules de calcul :

on calcule le ratio de réintégration : moyenne de la variable d'intérêt dans l'échantillon rapportée à la moyenne de la variable auxiliaire.

$$r = \frac{\bar{y}}{\bar{x}}$$

on calcule le total estimé  $\hat{T}$  de la variable d'intérêt (montant extrapolé) en appliquant le ratio de réintégration sur le total de la variable auxiliaire dans la population :

$$\hat{T} = rF$$

## SONDAGE ALEATOIRE STRATIFIE

### EXTRAPOLATION PAR LE NOMBRE D'INDIVIDUS DE LA POPULATION

#### Calcul de l'extrapolation

Notation :

$N$	Taille de la population
$T$	total variable d'intérêt dans la population
$\hat{T}$	Total estimé de la variable d'intérêt dans la population
$H$	nombre de strates
$h$	indice d'une strate : $h \in \{1, H\}$
$y$	variable d'intérêt
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon
$i$	indice d'un individu : $i \in \{1, n\}$

Formules de calcul :

**Pour chaque strate  $h$**  on calcule la moyenne de la variable d'intérêt dans l'échantillon. C'est le total de la variable d'intérêt dans l'échantillon divisé par le nombre d'individus de l'échantillon.

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_i}{n_h}$$

On calcule ensuite la moyenne pondérée de  $y$  sur la population

$$\bar{y}_{st} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h$$

le total estimé  $\hat{T}$  de la variable d'intérêt pour la population est égale au produit de la moyenne pondérée de la variable d'intérêt  $\bar{y}_{st}$  par l'effectif de la population.

$$\hat{T} = N \bar{y}_{st}$$

## SONDAGE ALEATOIRE STRATIFIE

### EXTRAPOLATION PAR LE RATIO SEPARÉ PAR STRATE

#### Calcul de l'extrapolation

Notation :

$N$	Taille de la population
$T$	total variable d'intérêt dans la population
$\hat{T}$	Total estimé de la variable d'intérêt dans la population
$F$	total variable auxiliaire dans la population
$H$	nombre de strates
$h$	indice d'une strate : $h \in \{1, H\}$
$F_h$	Total variable auxiliaire pour la strate $h$
$y$	variable d'intérêt
$x$	variable auxiliaire
$\bar{x}$	moyenne de $x$ dans l'échantillon
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon
$i$	indice d'un individu : $i \in \{1, n\}$

Formules de calcul :

**Pour chaque strate  $h$**  on calcule le ratio de réintégration : moyenne de la variable d'intérêt dans l'échantillon rapportée à la moyenne de la variable auxiliaire.

$$r_h = \frac{\bar{y}_h}{\bar{x}_h}$$

De même **pour chaque strate  $h$**  on calcule le total estimé  $\hat{T}_h$  de la variable d'intérêt (montant extrapolé) en appliquant le ratio de réintégration sur le total de la variable auxiliaire.

$$\hat{T}_h = r_h F_h$$

le total estimé  $\hat{T}$  de la variable d'intérêt pour la population (montant extrapolé) est égale à la somme des totaux estimés  $\hat{T}_h$  de la variable d'intérêt pour chaque strate  $h$  :

$$\hat{T} = \sum_{h=1}^H \hat{T}_h$$

## SONDAGE ALEATOIRE STRATIFIE

### EXTRAPOLATION PAR LE RATIO COMBINE

#### Calcul de l'extrapolation

Notation :

$N$	Taille de la population
$T$	total variable d'intérêt dans la population
$\hat{T}$	Total estimé de la variable d'intérêt dans la population
$F$	total variable auxiliaire dans la population
$H$	nombre de strates
$h$	indice d'une strate : $h \in \{1, H\}$
$y$	variable d'intérêt
$x$	variable auxiliaire
$\bar{x}$	moyenne de $x$ dans l'échantillon
$\bar{y}$	moyenne de $y$ dans l'échantillon
$n$	Taille de l'échantillon
$i$	indice d'un individu : $i \in \{1, n\}$

Formules de calcul :

**Pour chaque strate  $h$**  on calcule la moyenne de la variable d'intérêt dans l'échantillon.  $\bar{y}_h$

On calcule ensuite la moyenne pondérée par strate de la variable d'intérêt  $\bar{y}_{st}$

$$\bar{y}_{st} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h$$

**Pour chaque strate  $h$**  on calcule la moyenne de la variable auxiliaire dans l'échantillon.  $\bar{x}_h$  On calcule ensuite la moyenne pondérée par strate de la variable auxiliaire  $\bar{x}_{st}$

$$\bar{x}_{st} = \sum_{h=1}^H \frac{N_h}{N} \bar{x}_h$$

on calcule le ratio de réintégration : moyenne pondérée par strate de la variable d'intérêt dans l'échantillon rapportée à la moyenne pondérée par strate de la variable auxiliaire.

$$r = \frac{\bar{y}_{st}}{\bar{x}_{st}}$$

on calcule le total estimé  $\hat{T}$  de la variable d'intérêt (montant extrapolé) en appliquant le ratio de réintégration sur le total de la variable auxiliaire dans la population :

$$\hat{T} = rF$$



**SONDAGE ALEATOIRE STRATIFIE****STRATIFICATION PAR NUEE DYNAMIQUES**

Il s'agit de réaliser une partition de la population de telle sorte que la variance de la variable d'intérêt soit la plus faible possible à l'intérieur de chaque strate.

On utilise la méthode dite de regroupement autour de centres mobiles, appliquée à une seule variable auxiliaire.

**Algorithme**

- 1) Saisie du nombre de strates souhaité  $h$  et de la variable auxiliaire retenue  $x$ .
- 2) La population à stratifier est ordonnée selon la variable auxiliaire  $x$ .
- 3) on détermine  $h$   $P_i$  variables du programme tels que :  
 $i$  indice de la strate  
 $P_1 = \text{minimum de } x \text{ sur la population}$   
 $P_h = \text{maximum de } x \text{ sur la population}$   
les  $P_i$  restantes sont réparties de manière équidistante

**DEBUT DE LA BOUCLE**

- 4) on détermine  $h + 1$   $B_i$  variables du programme tels que :  
 $B_1 = \text{minimum de } x \text{ sur la population}$   
 $B_{h+1} = \text{maximum de } x \text{ sur la population}$   
les  $B_i$  restantes sont les bissectrices de  $[P_i; P_{i+1}]$
- 5) on recalcule les  $h$   $P_i$  variables du programme tels que :  
pour tout  $i$ ,  $P_i$  est la moyenne de  $x$  sur la strate déterminée par les bornes  $[B_i; B_{i+1}]$
- 6) si pour tout  $i$ ,  $P_i \text{ final} - P_i \text{ initial} > e$  refaire la boucle  
sinon sortir de la boucle

**FIN DE LA BOUCLE**

les  $h+1$   $B_i$  trouvées sont les bornes qui permettent de stratifier la population

**SONDAGE ALEATOIRE SIMPLE****TIRAGE SYSTEMATIQUE ORDONNE : ALGORITHME D'ARDILLY**

On calcule d'abord un pas de tirage :  $PAS = \frac{N}{n}$

( PAS n'est pas nécessairement un nombre entier )

On calcule le rang du premier individu :

X est un nombre réel au hasard entre 0 et 1  $X \in \mathfrak{R} / X \in [0,1[$

le rang est :  $rang = 1 + \text{int}(X.PAS)$

où int représente la partie entière d'un nombre positif.

On calcule le rang des individus pour les n-1 individus restants.

I est un indice de boucle entier variant de 1 à n-1

$I \in N / \forall I \in [1, n-1]$

$rang(I) = 1 + \text{int}[(X + I)PAS]$

## SONDAGE ALEATOIRE STRATIFIE

### TIRAGE SYSTEMATIQUE ORDONNE : ALGORITHME D'ARDILLY

L'algorithme utilisé est le même que celui pour le sondage aléatoire simple.

Il est dans ce cas appliqué strate par strate et non sur la population globale.

#### **Pour chaque strate h:**

On calcule d'abord un pas de tirage :  $PAS = \frac{N_h}{n_h}$

( PAS n'est pas nécessairement un nombre entier )

On calcule le rang du premier individu :

X est un nombre réel au hasard entre 0 et 1  $X \in \mathfrak{R} / X \in [0,1[$

le rang est :  $rang = 1 + \text{int}(X.PAS)$

où int représente la partie entière d'un nombre positif.

On calcule le rang des individus pour les  $n_h - 1$  individus restants.

I est un indice de boucle entier variant de 1 à  $n_h - 1$

$I \in N_h / \forall I \in [1, n_h - 1]$

$rang(I) = 1 + \text{int}[(X + I)PAS]$

**SONDAGE ALEATOIRE STRATIFIE****ALLOCATION DE NEYMAN**

Calcul du taux de sondage pour chaque strate  $h$  correspondant à la répartition optimale par strate des individus de l'échantillon global.

Formules de calcul :

Allocation de Neyman :

$$\frac{f_h^{optimum}}{f} = \frac{S_h}{\bar{S}}$$

avec :

écart type de la strate  $h$

écart type moyen des strates

$$\bar{S} = \sum_{h=1}^H \frac{N_h}{N} S_h$$