

Contrôles a posteriori d'enquêtes auprès des ménages

Exemple d'échantillonnage adapté en
fonction de la collecte

Les contrôles de collecte

- Donner des garanties de qualité de la collecte et de suivi des consignes
- Repérer le plus grand nombre d'anomalies possible et le plus tôt possible dans la collecte
- Sous contraintes de moyens et de rareté d'observation

Les contrôles de collecte

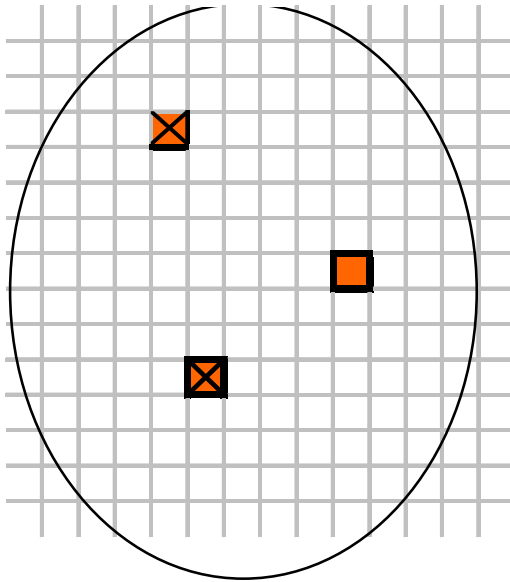
- Modalités des contrôles
 - Questionnaire adapté selon le statut de l'enquête (acceptée, refusée)
 - Enquête de contrôle par téléphone ou par courrier
- Deux exemples

	Saint Etienne (contrôle par téléphone principalement)	Grenoble (contrôle par courrier principalement)
Echantillon de l'enquête	6 500	10 400
Echantillon de contrôle	1 360	2 525
Taux de réponse au contrôle	63%	58%
<i>Dont enquêtés ayant refusé l'enquête</i>	<i>60%</i>	<i>40%</i>

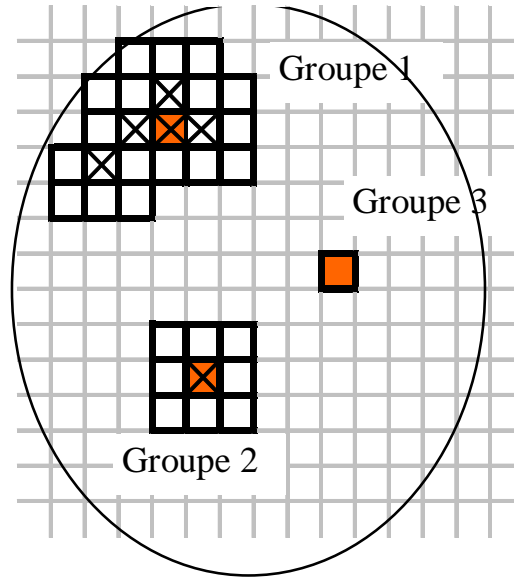
Stratégie d'échantillonnage

- Utiliser l'information collectée au fur et à mesure ...
- ... pour mieux cibler l'échantillon
- La sélection des enquêtes à contrôler devient fonction des résultats des précédents contrôles (adaptive sampling)

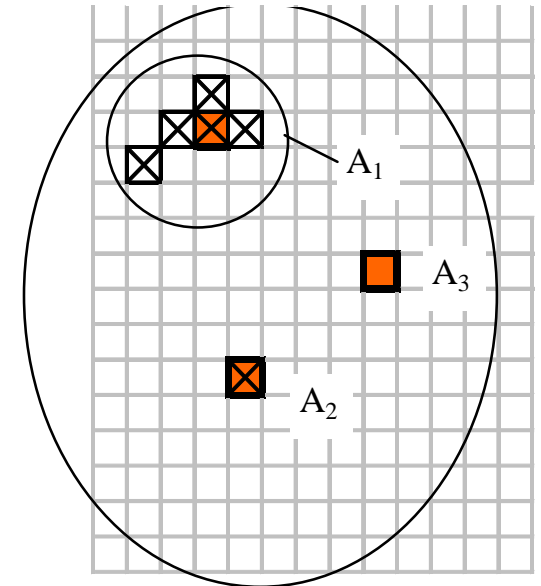
Adaptation de l'échantillonnage



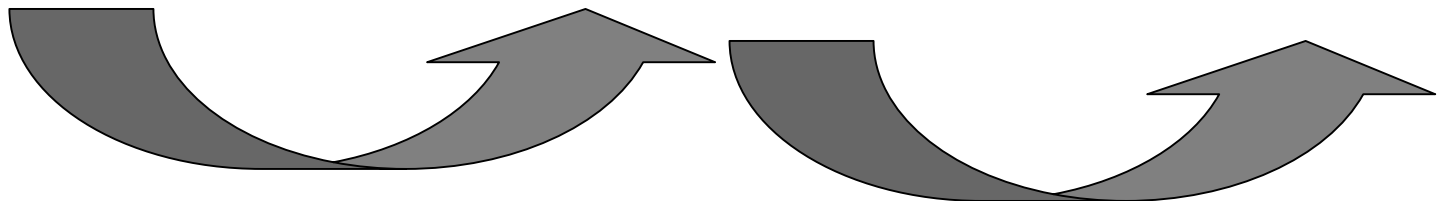
Echantillon initial



Groupes



Réseaux



Estimation du nombre d'anomalies

- On note :
 - N = taille de la population
 - m_i : le nombre d'unités du réseau i
 - a_i : le nombre d'unités des réseaux pour lesquels i est une unité marginale
 - n_1 : la taille de l'échantillon initial
- Calcul des probabilités d'inclusion
Le nombre total d'échantillon ne contenant pas i est :

$$C_{N - m_i - a_i}^{n_1}$$

La probabilité d'inclusion de i vaut :

$$\pi_i = 1 - \left[C_{N - m_i - a_i}^{n_1} \frac{1}{C_N^{n_1}} \right]$$

Estimation du nombre d'anomalies

On estime cette probabilité par une probabilité partielle:

$$\alpha_i = 1 - \left[C_{N-m_i}^{n_1} \frac{1}{C_N^{n_1}} \right]$$

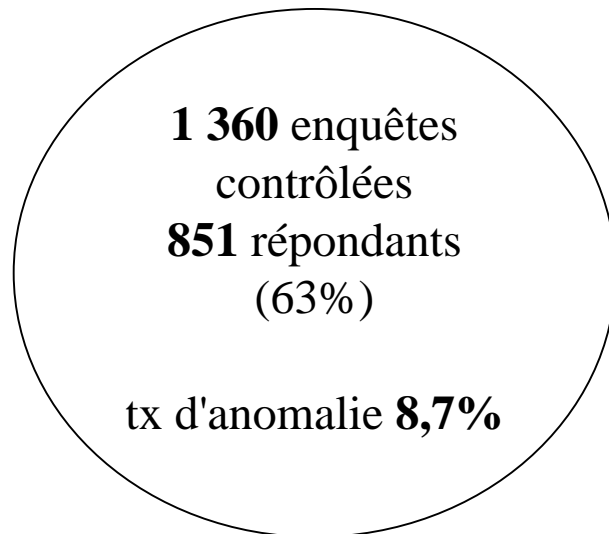
- Estimateur classique HT

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^k \frac{y_i}{\alpha_i}$$

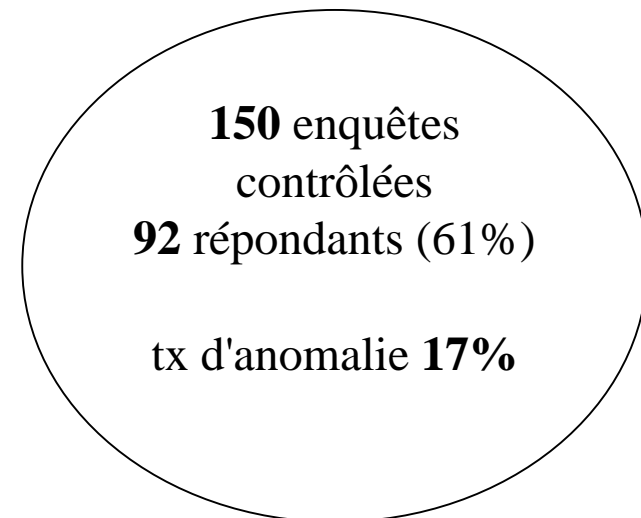
Enquête sur la région stéphanoise

- Plan de sondage : SAS
- Contrôles par téléphone

Echantillon initial



Echantillon supplémentaire



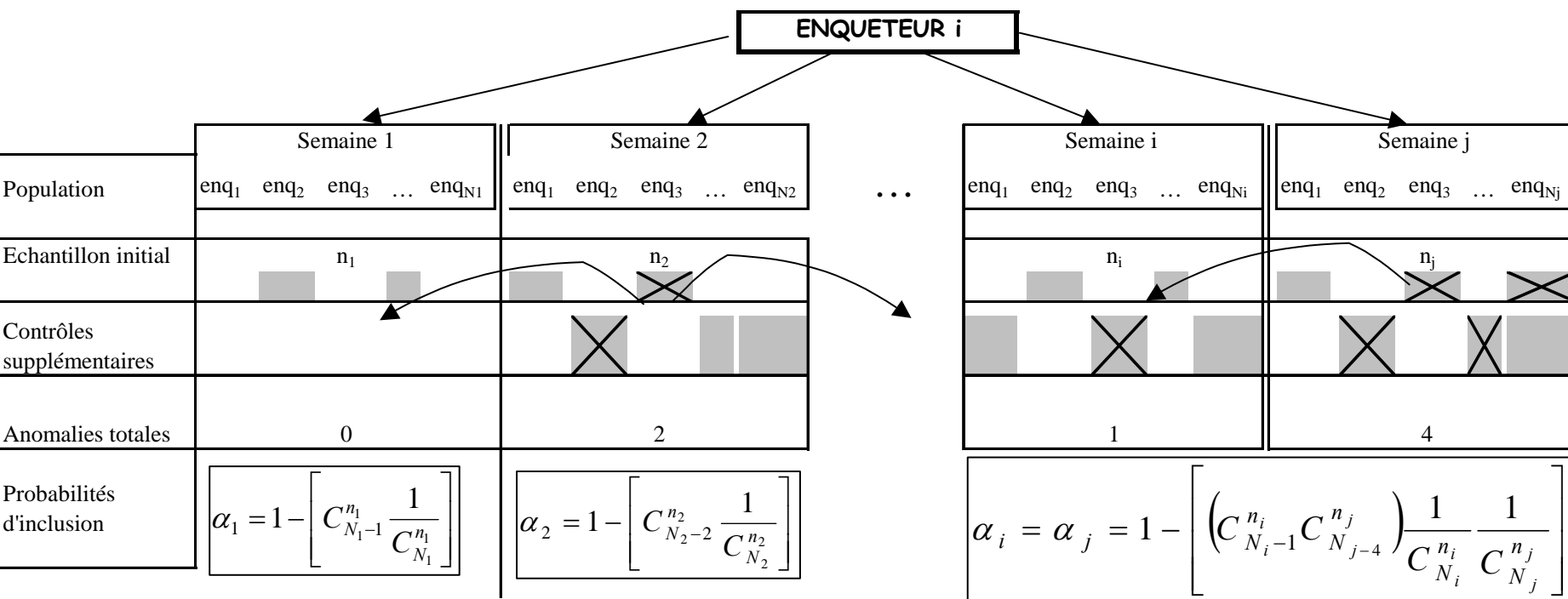
Taux global d'anomalies : 9,7%

(estimateur « adapté »)

Echantillon adapté et stratification

- Cibler d'avantage les contrôles de collecte en stratifiant préalablement la base de sondage
- Anomalie = $f(\text{temps, enquêteurs, terrain})$
- Intersection entre les strates et les voisinages

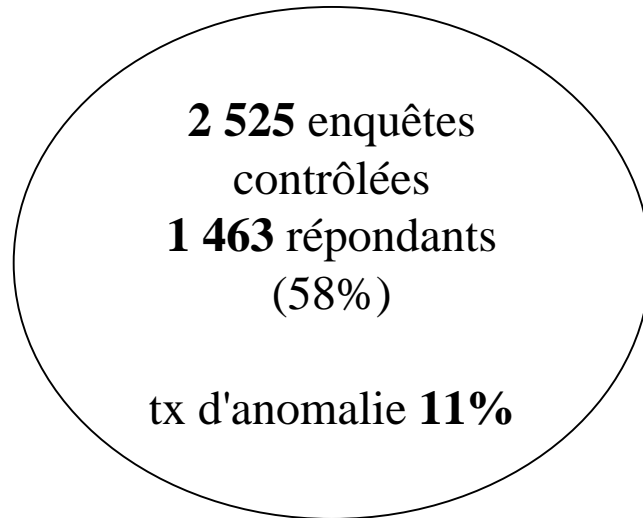
Echantillon adapté et stratification



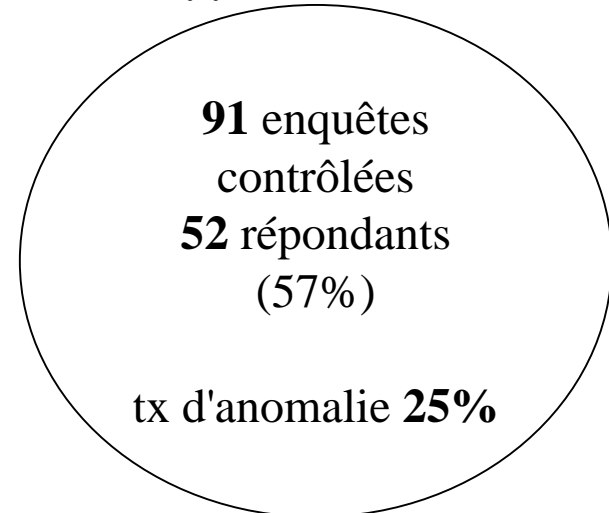
Enquête sur la région grenobloise

- Plan de sondage : SAS dans chaque strate
- Contrôles par courrier

Echantillon initial



Echantillon supplémentaire



Taux global d'anomalies : 12,1%

(estimateur « adapté »)

Synthèse des contrôles

- Stratégie de sélection efficace
 - Justifiée par une forte concentration des anomalies
 - Réduction des problèmes de collecte devenus rares en fin d'enquête
 - Charge de travail de contrôles raisonnable
- Estimation fiable du taux d'anomalies
- Conséquences sur les estimations des variables de l'enquête ?