

Intérêt et limites des estimations sur petits domaines

**HID « Petits domaines »,
une opération inhabituelle :**

- a. Une opération de « production »
- b. Visant l'ensemble des variables d'une enquête
- c. Recherchant une mise en œuvre simplissime



Les documents et outils disponibles

Un kit de calcul adapté aux zones d'étude

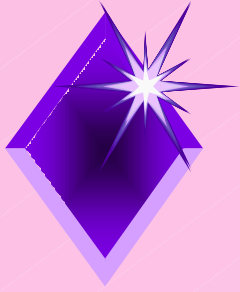
→ *Livré aux sept DR INSEE concernées*

**Un document à lucarne retraçant les travaux
du groupe « HID-Petits domaines »**

→ *Disponible auprès de la division EED*

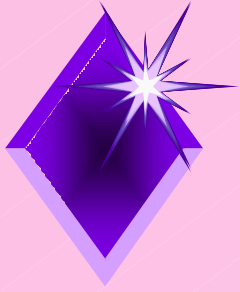
**Un volume de tableaux présentant les
estimations sur les huit zones d'étude**

→ *Insee-Résultats en cours d'impression*



Plan de l'exposé

- Fournir des estimations départementales (1)
- L'hypothèse de base (2-3)
- Biais et variance (4)
- La clé : de bons critères de stratification
Profiter du RP et de l'enquête VQS (5 à 8)
- Interprétation des résultats et discussion des enseignements (9)



(1) Pourquoi des estimations départementales ?

La demande locale

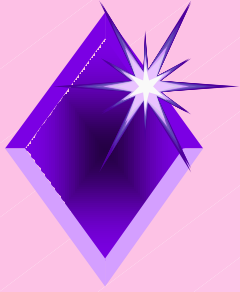
- le rôle essentiel des Conseils Généraux dans l'aide aux personnes âgées ou handicapées

L'insuffisance des enquêtes nationales

- des échantillons non représentatifs au niveau des départements

Un coût prohibitif des extensions locales

- 20 000 pour la France, versus 100 fois 2 000 pour les départements : **un coût multiplié par dix**



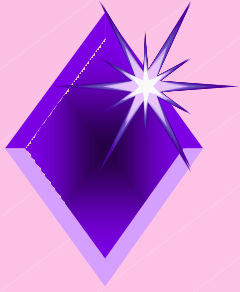
(2) Rappel de l'hypothèse classique de la méthode « petits domaines »

L'hypothèse de comportement homogène...

- elle suppose que les individus peuvent être rangés dans des groupes ayant un "**comportement moyen**" identique quelle que soit la zone géographique étudiée
- par "**comportement**", on entend ici la prévalence des divers aspects du handicap

...définit la méthode de travail

- 1. rechercher les critères ayant un bon "pouvoir explicatif" des prévalences*
- 2. rassembler des statistiques fournissant la répartition de la population aux niveaux national et local selon ces critères*



(3) L'estimateur petits domaines

L'estimateur direct (le "résultat brut")

Il fait intervenir pour la zone étudiée "Z"

- ❖ la prévalence du phénomène étudié dans chacun des groupes définis par les critères (Y_{zg})
- ❖ et la répartition des effectifs de la zone dans chaque groupe (P_{zg}/P_z)

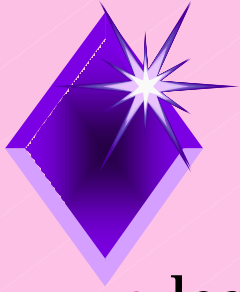
$$\bar{y}_z = \sum_{\text{groupe } g} \bar{y}_{zg} \frac{P_{zg}}{P_z}$$

L'estimateur "petits domaines"

La seule différence porte sur la prévalence

- ❖ on utilise celle de chacun des groupes au niveau national ($Y.g$)

$$\bar{y}_z = \sum_{\text{groupe } g} \bar{y}_{.g} \frac{P_{zg}}{P_z}$$



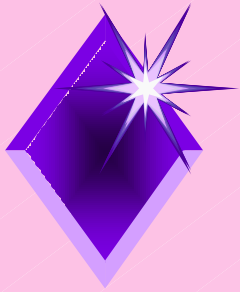
(4) Biais et variance (a)

- les “Y.g” nationaux sont toujours différents des “Yzg”. Ceci constitue un **Biais** de l’estimation
- mais il sont estimés sur la totalité de l’échantillon national. **L’intervalle de confiance** de l’estimation “petits domaines” est donc plus resserré.

Contrôle sur le cas de l’Hérault (où on a une extension HID)

L’estimateur local direct est établi sur un effectif important (1500 réponses) alors que sur un département standard, on a en moyenne 170 réponses.

Le gain sur l’intervalle de confiance est donc fortement minimisé, alors que le biais est inchangé



(4) Réduction de l'aléa mais introduction d'un biais (b)

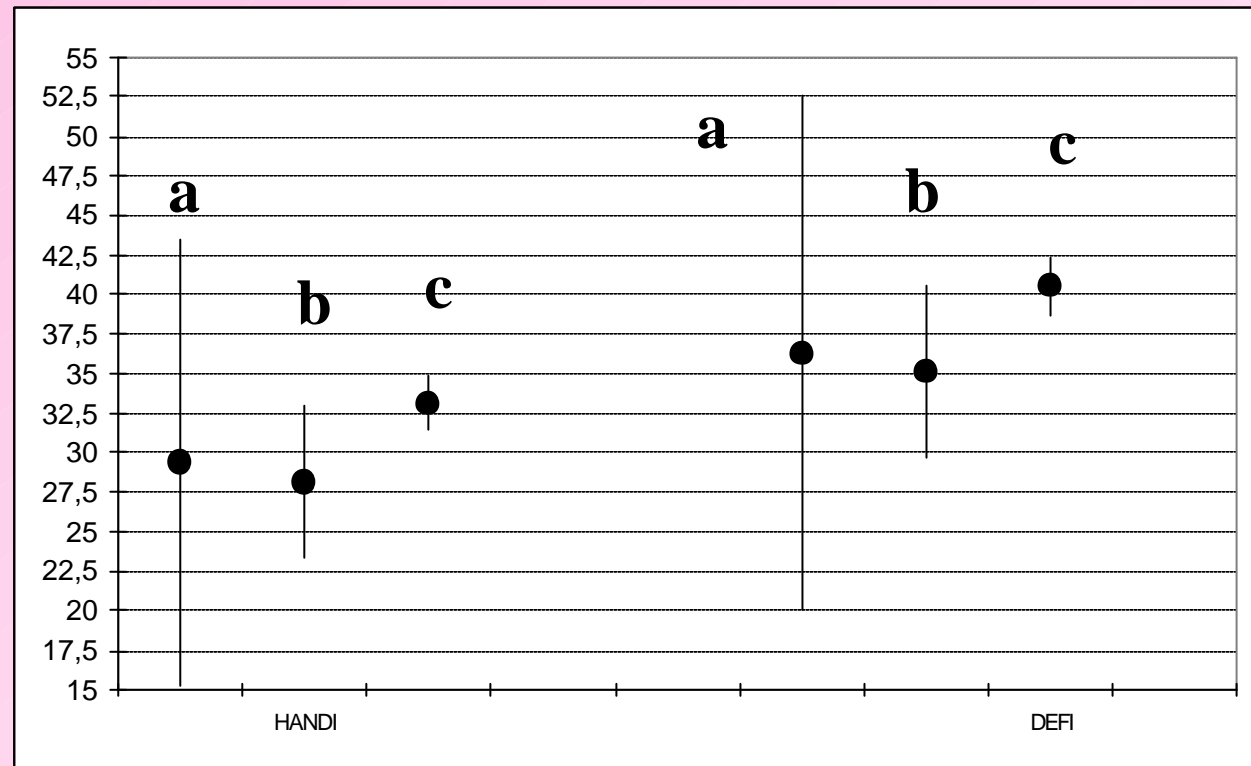
3 estimateurs :

a) l'estimateur direct sur un l'échantillon HID de l'Hérault sans extension

b) l'estimateur direct sur l'échantillon HID de l'Hérault

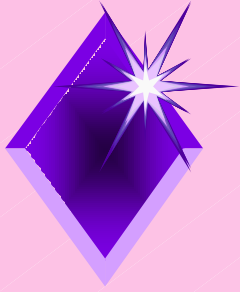
(calés sur la structure VQS de l'Hérault)

c) l'estimateur "petits domaines"



2 variables d'intérêt : Handi

DEFI



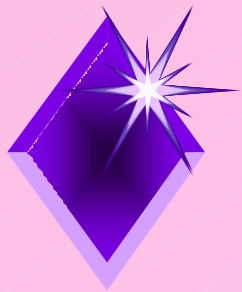
(5) La clé : disposer de bons critères de "stratification"

L'enjeu

- Mieux on expliquera le comportement, plus on **limitera le biais de l'estimation « petits domaines »**

Une double limite

- **La disposition** des critères voulus **au niveau local**
- **La taille de l'échantillon national**
 - Qui réduit drastiquement le nombre de croisements admissible



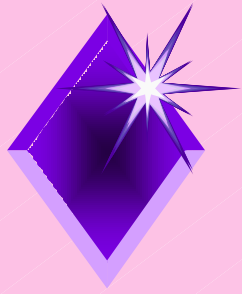
(6) *Un atout : la proximité du recensement*

L'apport du recensement

- la proximité de l'enquête et du recensement de 1999 assure la disposition de données socio-démographiques variées à tout niveau géographique : sexe, âge, catégorie sociale, niveau d'études, zone d'habitat...
- néanmoins ce sont toutes des variables indirectement corrélées au handicap

... et celui de son complément VQS

- 1. on a recherché un indicateur directement lié au handicap*
- 2. faute de donnée spatialement homogène, on en a construit une à l'aide d'une enquête supplémentaire au recensement : Vie Quotidienne et Santé*



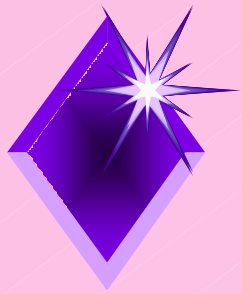
(7) La recherche des critères de "stratification"

La modélisation

- ❖ On teste par des techniques de régression (surtout logistiques) le “pouvoir explicatif” de différents croisements de critères sur différents aspects relevant du handicap
- ❖ Les critères sont choisis sous trois conditions :
 - avoir du “sens” (être interprétables)
 - avoir un effet (mesuré par les modèles testés)
 - ... être disponibles

On a retenu

- ❖ En premier l'âge et le groupe VQS, en second le sexe et la tranche d'unité urbaine, soient 180 « strates » ... qu'on a regroupées en 52



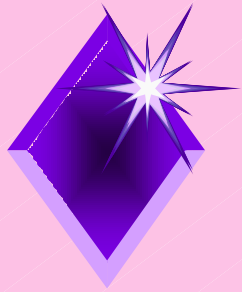
(8) Un mot sur la préenquête VQS et ses extensions locales

VQS : des objectifs nationaux...

- ❑ Pour tirer un échantillon comportant une forte proportion de personnes handicapées, on a réalisé une préenquête auprès de 400 000 personnes.
- ❑ Elle a permis de définir une variable “sévérité probable de handicap” (le groupe VQS), servant de base à un tirage à probabilités inégales,
- ❑ et fournissant un indicateur fortement corrélé au handicap

... et locaux

- 1. Cet indicateur fait naturellement partie des critères “explicatifs” retenus*
- 2. et les extensions locales de la préenquête (20000 par zone) ont permis de disposer de ce critère pour les départements demandeurs*



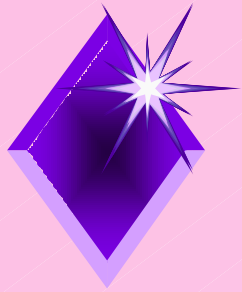
(9) L'interprétation des estimateurs locaux (a)

1. Avec les critères socio-démographiques, on corrige l'estimateur France Entière des particularités de structure locale

❖ *Restent les spécificités de comportement : même la prise en compte du milieu social (en 6 groupes) passerait à côté du problème de la silicose, qui est un phénomène localisé*

2. Avec la prise en compte du « groupe VQS », on prend partiellement en compte les spécificités locales de comportements

❖ *Parce que la variable utilisée peut être considérée comme un indicateur synthétique relevant du champ de l'enquête*



(9) L'interprétation des estimateurs locaux (b)

3. Mais le champ du handicap est multi-dimensionnel, et ne peut être résumé en un seul indicateur, quel qu'il soit. Ceci vaut pour les spécificités locales en matière de handicap

❖ *La variable « groupe VQS » n'est donc au mieux qu'une approche moyenne de la situation locale.*

4. En outre le « groupe VQS » ressort davantage du subjectif que de l'objectif,

❖ *Les modalités de l'enquête VQS accentuent fortement le caractère déclaratif ou subjectif de la variable construite*

❖ *Or dans le champ de la santé, la distance entre déclaration et situation objective peut s'avérer assez variable selon les spécificités culturelles (cf. l'expérience de Panel Européen sur la santé)*