

Conclusion

Séminaire sur les données administratives



DEUX GRANDES COMPOSANTES :

- Méthodologie d'acquisition des données
- Méthodologie de production de statistiques et d'analyses statistiques

PRISE EN COMPTE DE L'ENSEMBLE DES ERREURS

- Estimation de la distribution des erreurs (écarts) en lien avec les mécanismes aléatoires (biais-variance des estimateurs)



PRISE EN COMPTE DE L'ENSEMBLE DES ERREURS

- A priori, les données administratives conduisent à des estimateurs sans variance mais avec des biais relativement importants
- Nouveaux types d'erreurs : erreur d'appariement, erreur de géolocalisation, erreur de couverture, ...

SÉCURISER ET SIMPLIFIER LA GESTION DE LA CONFIDENTIALITÉ DES DONNÉES LOCALES

- Données carroyées ou infra-communales par exemple

REPENSER LE CONCEPT DE LOCALISATION POUR UN INDIVIDU

VERS DES MÉTHODES HYBRIDES (OU MULTI SOURCES) ?

- On parle aussi d'intégration de données
- On a déjà des cas d'usage :
 - Imputations par appariement, ERFS (EEC et données F&S)
 - Utilisation d'enquêtes de contrôle pour les méthodes DSE (capture-recapture) : exemple des EAR pour Résil
- On peut aller plus loin ?
 - Poids dans les sources issues de données Administratives ?

BASE DE SONDAGE

- **Passer de Fidéli à Résil constitue une évolution positive**
 - Champ plus large
 - Plus de données de contact des individus
 - Plus de variables auxiliaires
- **Défis associés :**
 - Gérer les valeurs manquantes dans la base de sondage (variables filtres pour le champ et variables de plan)
 - Se prémunir des évolutions ou disparitions des sources principales (cf TH, GMBI)

REDRESSEMENT

- Pour l'essentiel, le calage des enquêtes ménages s'appuie sur des marges estimées à partir de l'EEC

LE DÉFI

- Obtenir un univers de référence corrigé :
 - Des erreurs de couverture
 - Des valeurs manquantes (pour les variables de calage et de filtrage de champ)

Merci à tous et bonne fin de journée

insee.fr

