



## Internet est-il l'avenir des enquêtes Génération ?

**Journées de Méthodologie Statistique**

2 avril 2015

*Christophe Dzikowski (Céreq)*

*Christophe Barret (Céreq)*

# Les enquêtes Génération

- Enquête sur l'insertion professionnelle des jeunes à l'issue du système éducatif
- Constitution d'une base de sondage de jeunes présumés sortants auprès des établissements de formation initiale
- Enquête téléphonique (CATI) à 3, 5 et 7 ans après la sortie
- Entre 20000 et 60000 questionnaires selon les Générations
- **Nécessité de rénover le dispositif après 2017**
  - ↳ une des pistes basculement sur une collecte multimode internet (Cawi) et téléphone (Cati)

## Enquête 2013 auprès de la Génération 2010

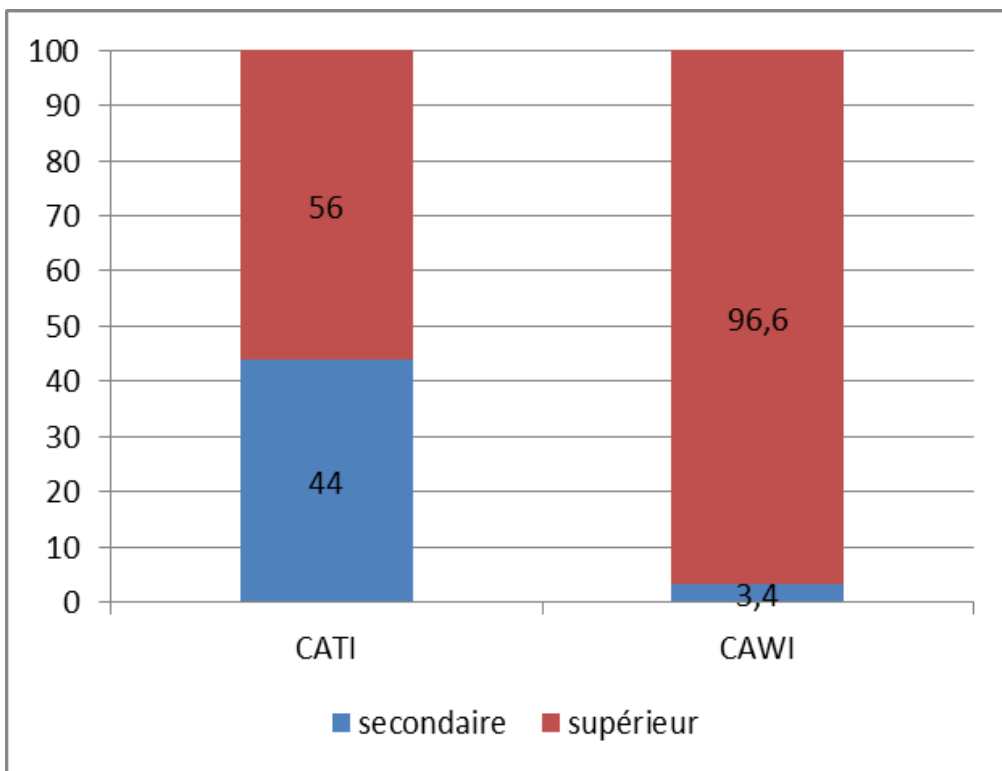
- Deux échantillons disjoints
- Deux collectes monomodes distinctes en parallèle
- Questionnaires quasiment identiques « copie miroir »
  - Quelques reformulations pour internet
  - Ajout de consignes pour certaines questions complexes et pour respecter certaines conventions
- Pour internet, envoi d'un mail-avis suivi de 10 relances mails

# De nombreuses difficultés de collecte

- Base de sondage peu adaptée à la collecte par internet
    - Notamment pas de mail pour les bases d'élèves issues du rectorat
    - Au final, seulement **¼ d'email disponible**
    - Tentative d'enrichissement des emails via un prestataire non concluante
  - Un envoi de mail en masse et un taux de connexion décevants
  - Des abandons en cours d'enquête plus nombreux
- ↳ 15% de répondants parmi ceux qui disposaient d'un mail

# Traitement préalable des données CAWI

Structure des répondants selon le niveau de sortie et le mode de collecte, poids initiaux



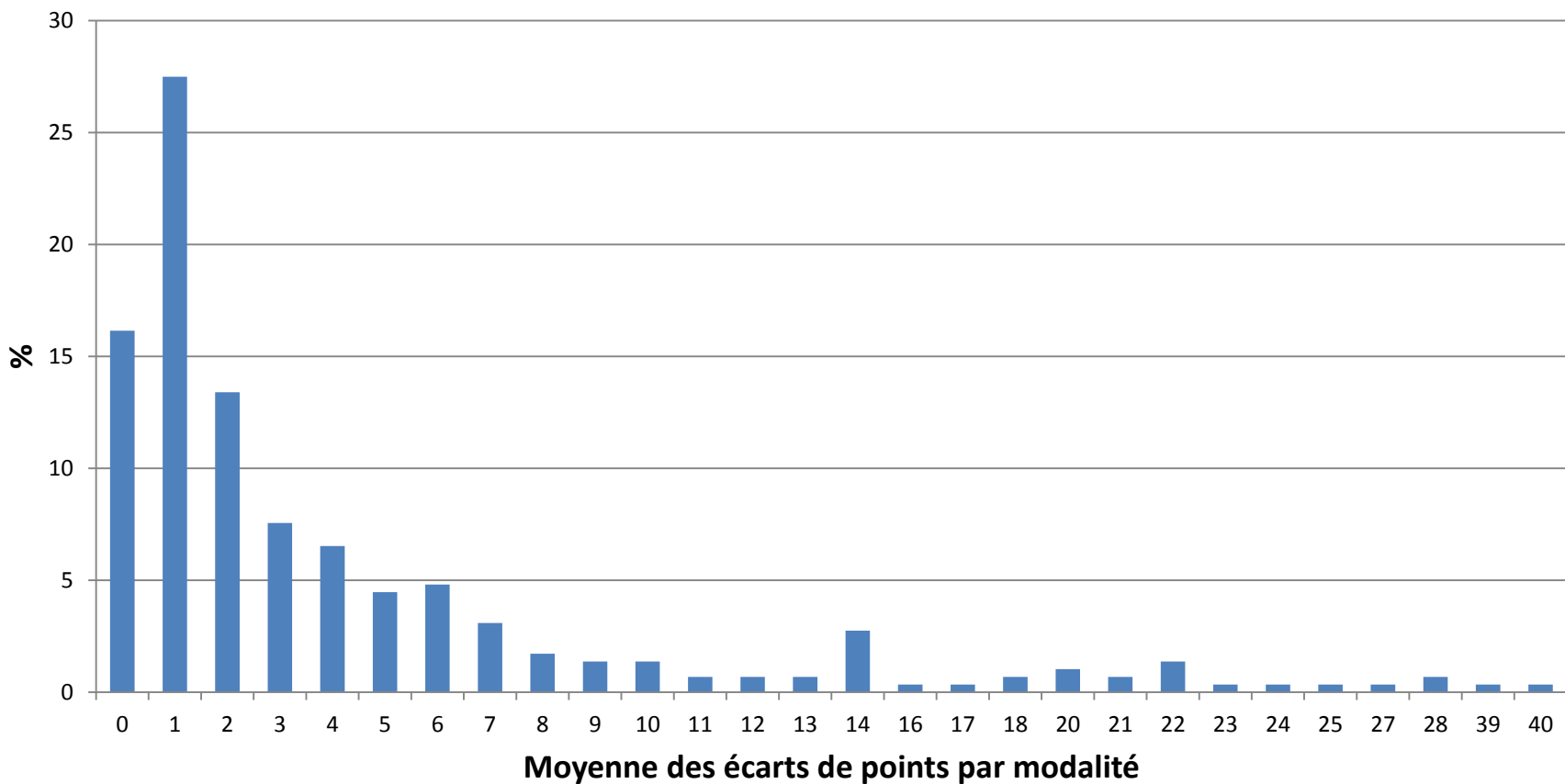
- Restriction sur une sous population du supérieur (4 096 observations)
- Protocole de modélisation de la non réponse et de calage quasi identique entre CATI et CAWI

## Exemple avec la variable situation à la date d'enquête

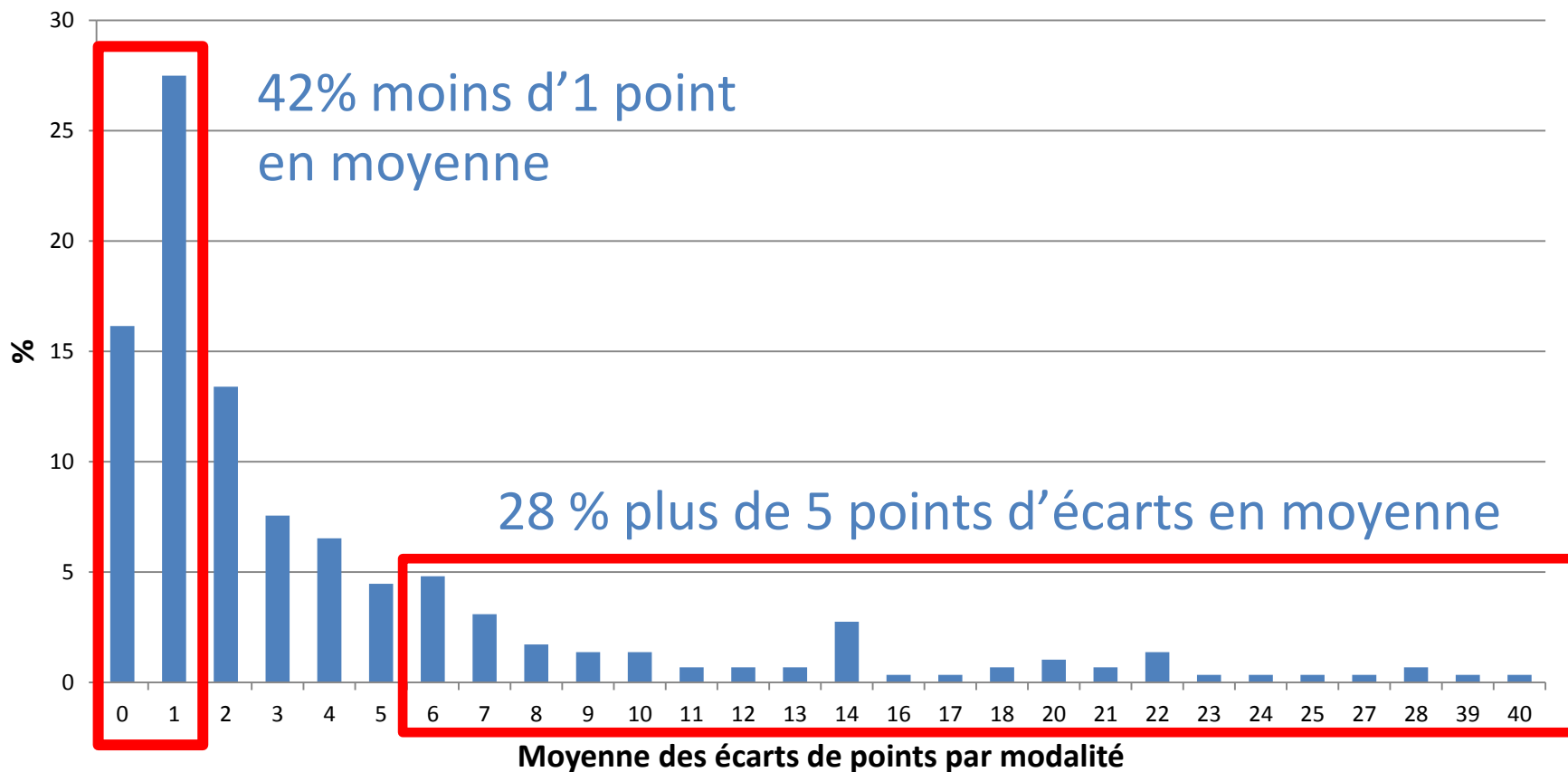
Situation à la date d'enquête	CAWI	CATI
emploi	76,5	75,9
chômage	11,9	12,5
inactivité	3,8	2,7
formation	1,2	1,7
reprise d'étude	6,4	7,2

- Total des écarts de points =  $|76,5 - 75,9| + |11,9 - 12,5| + |3,8 - 2,7| + |1,2 - 1,7| + |6,4 - 7,2| = 3,6$
- Moyenne des écarts de points par modalité = 0,72

# Structure des différences



# Structure des différences





# Après la prise en compte de la variance

- Estimation de la variance des proportions par calcul direct
  - En prenant en compte :
    - La variance du tirage équilibré
    - La variance due à la non réponse
    - En prenant en compte le calage
    - Appliquée à la linéarisée d'une proportion
  - Un peu plus de 40% des variables présentent une égalité statistique pour chacune des modalités
  - 30 % des variables présentent aucune égalité sur aucune des modalités
  - Égalité statistique partielle pour le reste des variables
- ↪ Les différences significatives s'expliquent par des **effets de mesure** et/ou de **sélection** importants

# Exemples d'effet de mesure a priori

Raisons d'arrêt des études (QCM à 8 modalités)	%CAWI (IC 95%)	%CATI (IC 95%)
Parce que vous aviez atteint le niveau de formation que vous souhaitiez (ATNIVO)	43(+1,1)	60(+0,7)

- Perception différente d'une QCM à l'oral et à l'écrit ?
- Un effet de désirabilité sociale plus important par téléphone ?

- **Objectif : distinguer les effets de sélection des effets de mesure**
- Méthode habituellement utilisée dans le cadre d'évaluation de politiques publiques (D. Fougère)
- Pour mesurer l'effet d'une politique publique, l'idéal est de procéder à une expérimentation avec une affectation aléatoire des bénéficiaires du dispositif pour éviter les problèmes de sélection.
- Sans expérimentation aléatoire préalable, le matching propose de contrôler l'effet de sélection entre les bénéficiaires et les non-bénéficiaires

# Le matching... pour distinguer les effets

- Dans notre cas, le traitement  $T$  est le fait de répondre par internet

$$T = \begin{cases} 1 & \text{pour les répondants au Cawi} \\ 0 & \text{pour les répondants au Cati} \end{cases}$$

- La variable d'intérêt  $Y$  peut s'écrire avec 2 variables latentes (observées partiellement)

$$Y = T \times Y_1 + (1 - T) \times Y_0$$

- La méthode repose sur l'hypothèse que conditionnellement aux variables qui expliquent la sélection, les variables latentes sont indépendantes du traitement
- Dans ce cas, il y a également indépendance par rapport à la probabilité de bénéficier du traitement (le score de propension)

$$(Y_0 \perp T|X) \implies (Y_0 \perp T|P(X)) \quad (\text{avec } P(X) = \Pr(T = 1|X))$$

# Le matching... en pratique

- 2 groupes : groupe répondant par CAWI, et le groupe de contrôle CATI
  - Pool de données CATI et CAWI avec partage des poids
- Modéliser les effets de sélection à partir de la modélisation du fait de répondre par internet (vs répondre par téléphone)
- Pour chaque répondant CAWI, on cherche des contrefactuels parmi les répondants CATI qui leur « ressemblent » (A l'aide des probabilités estimées de répondre au CAWI)
- L'effet de mesure est donné par la différence observée entre les statistiques observées sur le groupe des répondants CAWI et le groupe de leur contrefactuel du CATI (noté Cati \*).

# Le matching... application

- Exemple sur ATNIVO

Raisons d'arrêt des études (QCM à 8 modalités)	CAWI	CATI	Écart brut (Cawi-Cati)	Ecart avec prise en compte de l'effet de sélection (Cawi-Cati*)
Parce que vous aviez atteint le niveau de formation que vous souhaitiez (ATNIVO)	43%	60%	<b>-17 points</b>	<b>-20 points</b>

- La prise en compte de la sélection aggrave l'écart observé
- Mais le phénomène majeur reste un problème de mesure

- Test du balancing score confirme l'hypothèse d'indépendance

# Les limites de la méthode

- La distinction des effets de sélection et de mesure est délicate
- Elle dépend de la manière dont est spécifiée le modèle de sélection
- Si le modèle est trop ajusté on n'arrive plus à définir le groupe de contrefactuel

## Objectifs

- Améliorer le protocole de contact  
→ augmenter les taux de réponse
- Améliorer l'ergonomie du questionnaire  
→ réduire les abandons en cours de collecte  
→ réduire l'effet de mesure (generalized design)
- Améliorer la distinction des effets de mesure et de sélection
  - Méthode A : ajout des questions pour mieux définir la propension à répondre à internet
  - Méthode B : mise en place d'un échantillon embarqué avec affectation aléatoire du mode de collecte



# L'échantillon embarqué

- En reprenant l'analogie avec l'évaluation des PP, l'idée est de mettre en place une expérimentation aléatoire
- Un échantillon multimode « classique » (Cawi en majeur et Cati en rattrapage) (4444 individus)
- Un échantillon embarqué avec affectation aléatoire des modes de collecte (1000 individus)
  - L'idée est de multiplier les moyens de contacts
  - Une fois le contact établi (qu'il soit actif ou passif), on demande à l'enquêté s'il accepte de répondre à l'enquête.
  - S'il accepte on lui attribue aléatoirement un mode de collecte
  - **En théorie, protocole « idéal » pour gommer les effets de sélection.**
  - **En pratique, difficulté à terminer l'enquête (malgré l'acceptation de répondre) notamment lorsque l'on imposera de changer de mode.**

# L'agrégation en présence d'effet de mesure

- Dans un contexte multimode
  - L'effet de sélection est une chance → augmentation de la couverture
  - L'effet de mesure est un problème → comment agréger les 2 modes ?
- Hypothèse préalable :
  - on a réussi à identifier les variables qui posent un problème de mesure
  - Et, dans la plupart des cas, on parvient à quantifier l'effet de mesure (et l'effet de sélection)
- Nécessite de définir un mode référence
  - Comment le choisir ? Majoritaire, historique, « qualité », ...
  - Globalement ou variable par variable

# L'agrégation en présence d'effet de mesure

- En ne faisant rien !
- Calage
- Omission d'un mode de collecte puis imputation
- Transformation de réponse

# Conclusion : Internet est-il l'avenir des enquêtes Génération ?

- De nombreuses difficultés dans la collecte web...  
... mais qui paraissent surmontables
- L'enjeu des années à venir
  - Définir un Cawi et un Cati qui limitent les effets de mesure (problème du calendrier, des consignes et conventions...)
  - Malgré nos efforts des biais de mesure subsisteront
    - Soit on l'anticipe et on abandonne ces questions
    - Soit on trouve une méthode satisfaisante pour contourner cette difficulté (matching, échantillon embarqué puis agrégation des données)
- Un gage de modernité notamment pour la cible : les jeunes !...  
...Quitte à accepter de repenser en partie notre dispositif



## Internet est-il l'avenir des enquêtes Génération ?

**Journées de Méthodologie Statistique**

2 avril 2015

*Christophe Dzikowski (Céreq)*

*Christophe Barret (Céreq)*