

Écart dans la déclaration de la race et de l'origine hispanique entre les différentes sources de dossiers administratifs : Exploration de méthodes pour attribuer les réponses

Sharon Ennis, Sonya Rastogi et James Noon
U.S. Census Bureau

Symposium international sur les questions de méthodologie 2016 de
Statistique Canada
22 mars 2016

Avertissement : Cette communication est diffusée en vue d'informer les parties intéressées des travaux de recherche en cours et de favoriser l'examen de ces travaux. Les opinions exprimées sont celles des auteurs et ne représentent pas forcément celles du U.S. Census Bureau.

Aperçu

- Objectifs de la recherche
- Contexte
- Règles opérationnelles utilisées pour attribuer les réponses pour la race et l'origine hispanique dans les dossiers administratifs (DA)
- Analyse de régression logistique multinomiale
- Conclusion et prochaines étapes

Objectifs de la recherche

- Utiliser des règles opérationnelles pour l'attribution d'une réponse pour l'origine hispanique et d'une réponse pour la race lorsque ces données diffèrent entre les sources de DA
- Évaluer les caractéristiques des personnes dont les réponses pour l'origine hispanique et la race correspondent aux données du recensement, ne correspondent pas aux données du recensement ou sont manquantes dans les DA

Contexte

- Recherche sur la qualité des données comparant les réponses pour la race et l'origine hispanique des DA aux réponses du recensement
 - Recherche antérieure – DA de mauvaise qualité sur la race et l'origine hispanique pour les minorités
- Recherche sur la fluidité raciale
 - Recherche antérieure – Les changements de réponse sur la race sont courants parmi les Indiens de l'Amérique du Nord et les Autochtones de l'Alaska (IANAA), les Autochtones d'Hawaï et des Îles du Pacifiques (AHIP), les populations multiraciales et les populations hispaniques

Données

- Données du Recensement de 2010 (telles que déclarées)

Race dans les DA

- Trois fichiers différents du logement et développement urbain (HUD)
- Services de santé aux Indiens (IHS)
- Assurance-maladie (MEDB)
- Système d'identification numérique (Numident)
- Dossiers précédents du recensement
- Aide temporaire pour les familles dans le besoin (TANF)

Origine hispanique dans les DA

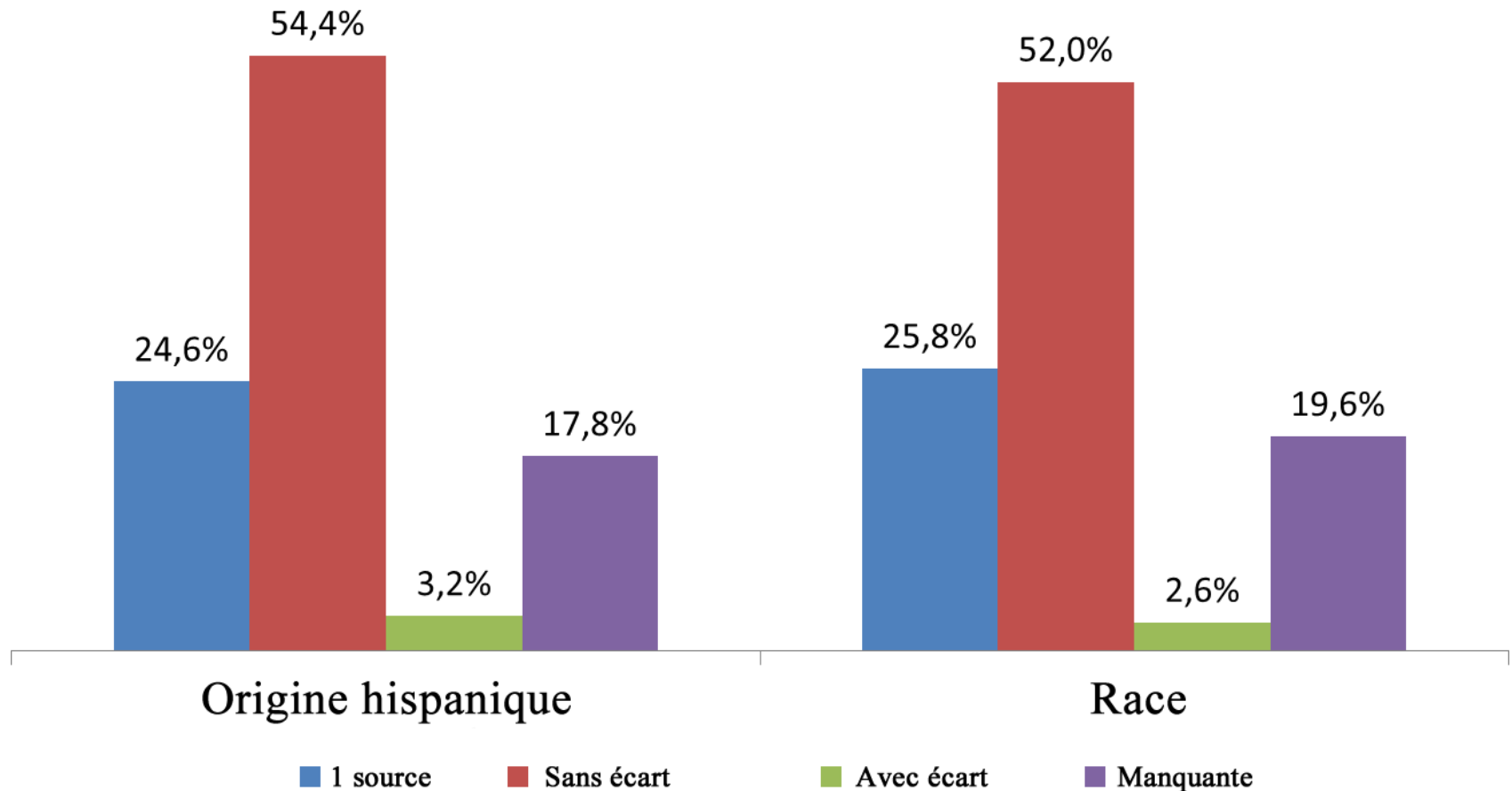
- Tous les fichiers sont utilisés pour la race dans les DA
- Système d'information statistique Medicaid (MSIS)
- Programme supplémentaire d'aide à la nutrition du Texas (SNAP)
- Quatre fichiers de tiers

Méthodologie

- Couplage d'enregistrements – Des identificateurs uniques, appelés clés d'identification protégées (CIP), ont servi à coupler les personnes des DA et du Recensement de 2010
- Taux d'appariement des réponses calculés de la façon suivante :

$$\frac{\text{Personnes ayant des réponses correspondantes dans les DA et dans le Recensement de 2010}}{\text{Personnes comprises dans le Recensement de 2010 pour lesquelles les DA fournissent des données démographiques}}$$

Pourcentage d'enregistrements comportant un écart entre les sources de DA



Méthode d'attribution des réponses pour l'origine hispanique dans les DA

Cas n°	Source de DA 1	Source de DA 2	Source de DA 3	Source de DA 4	Origine hispanique dans les DA
1	Non hispanique	Non hispanique	Non hispanique	Non hispanique	Non hispanique
2	Manquante	Non hispanique	Manquante	Manquante	Non hispanique
3	Hispanique	Hispanique	Manquante	Hispanique	Hispanique
4	Manquante	Hispanique	Manquante	Manquante	Hispanique
5	Hispanique	Non hispanique	Manquante	Hispanique	Hispanique
6	Non hispanique	Non hispanique	Non hispanique	Hispanique	Hispanique
7	Manquante	Manquante	Manquante	Manquante	Manquante

Méthode d'attribution des réponses pour la race dans les DA

Cas n°	Source de DA 1	Source de DA 2	Source de DA 3	Source de DA 4	Race dans les DA
1	Manquante	Blanche seulement	Blanche seulement	Blanche seulement	Blanche seulement
2	Noire seulement	Manquante	Noire seulement	Noire seulement	Noire seulement
3	AHIP seulement	Blanche seulement	Manquante	AHIP seulement	AHIP seulement
4	AHIP seulement	IANAA seulement	Blanche seulement	Manquante	AHIP seulement
5	IANAA seulement	Blanche seulement	Manquante	IANAA seulement	IANAA seulement
6	Asiatique seulement	IANAA seulement	Manquante	IANAA seulement	IANAA seulement
7	Asiatique seulement	Manquante	Noire seulement	Asiatique seulement	Asiatique seulement
8	Manquante	Manquante	Manquante	Manquante	Manquante

Autre méthode utilisée pour attribuer l'origine hispanique dans les DA

- Origine hispanique dans les DA, méthode 2
 - Réponse pour l'origine hispanique attribuée à partir des dossiers du recensement précédent. Si aucune réponse dans les dossiers du recensement précédent, l'origine hispanique est attribuée si elle est présente dans une autre source de DA.

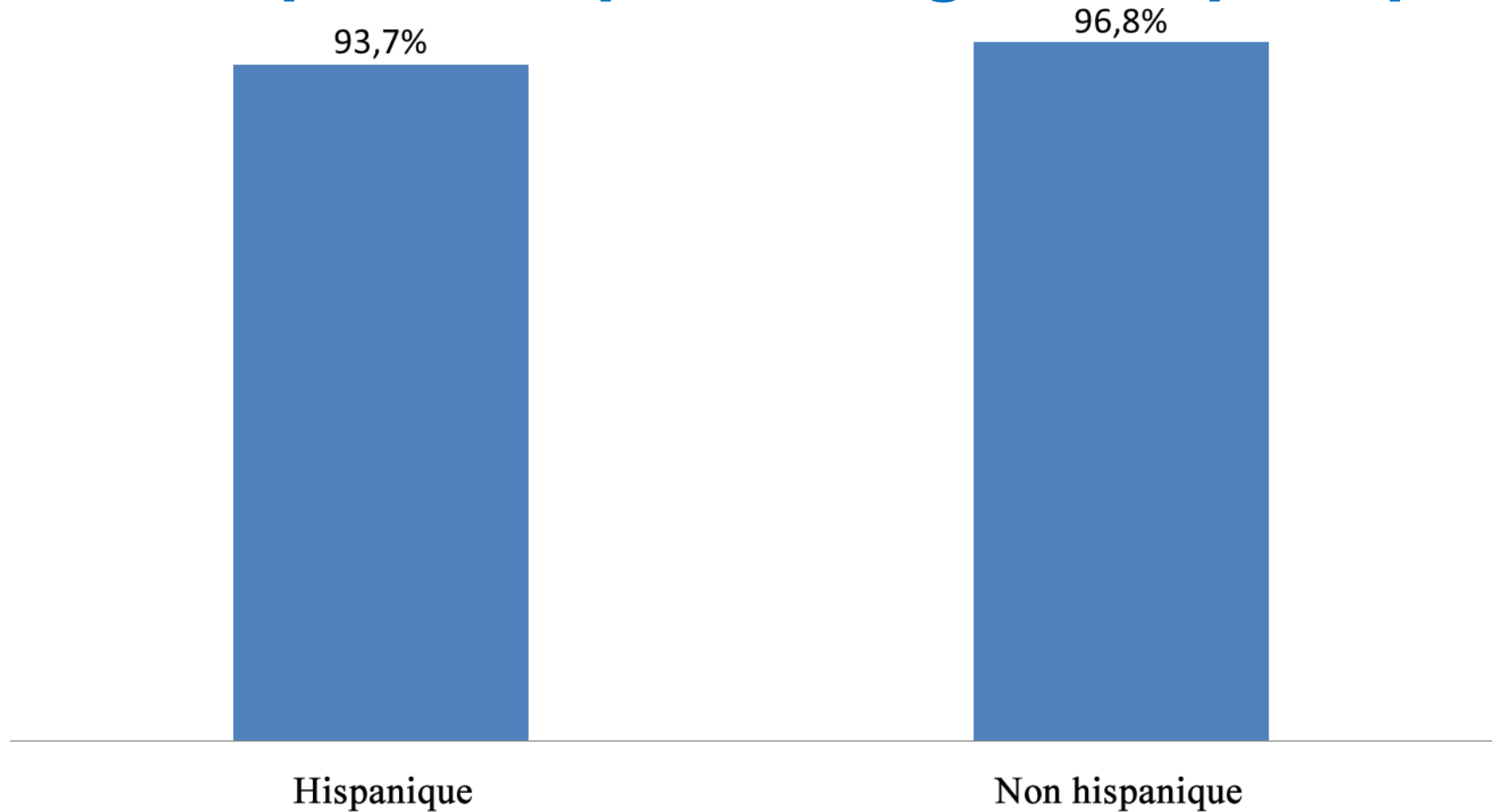
Autres méthodes utilisées pour attribuer la race dans les DA

- Race dans les DA, méthode 2
 - Similaire à la méthode 1, mais préférence accordée à deux races ou plus en premier.
- Race dans les DA, méthode 3
 - Race attribuée sur la base de la race la plus fréquemment déclarée parmi les sources de DA. En l'absence de race la plus fréquente, la race a été attribuée selon la méthode 1.
- Race dans les DA, méthode 4
 - Race attribuée sur la base de la race la plus fréquemment déclarée. En l'absence de race la plus fréquente, la race a été attribuée selon la méthode 2.

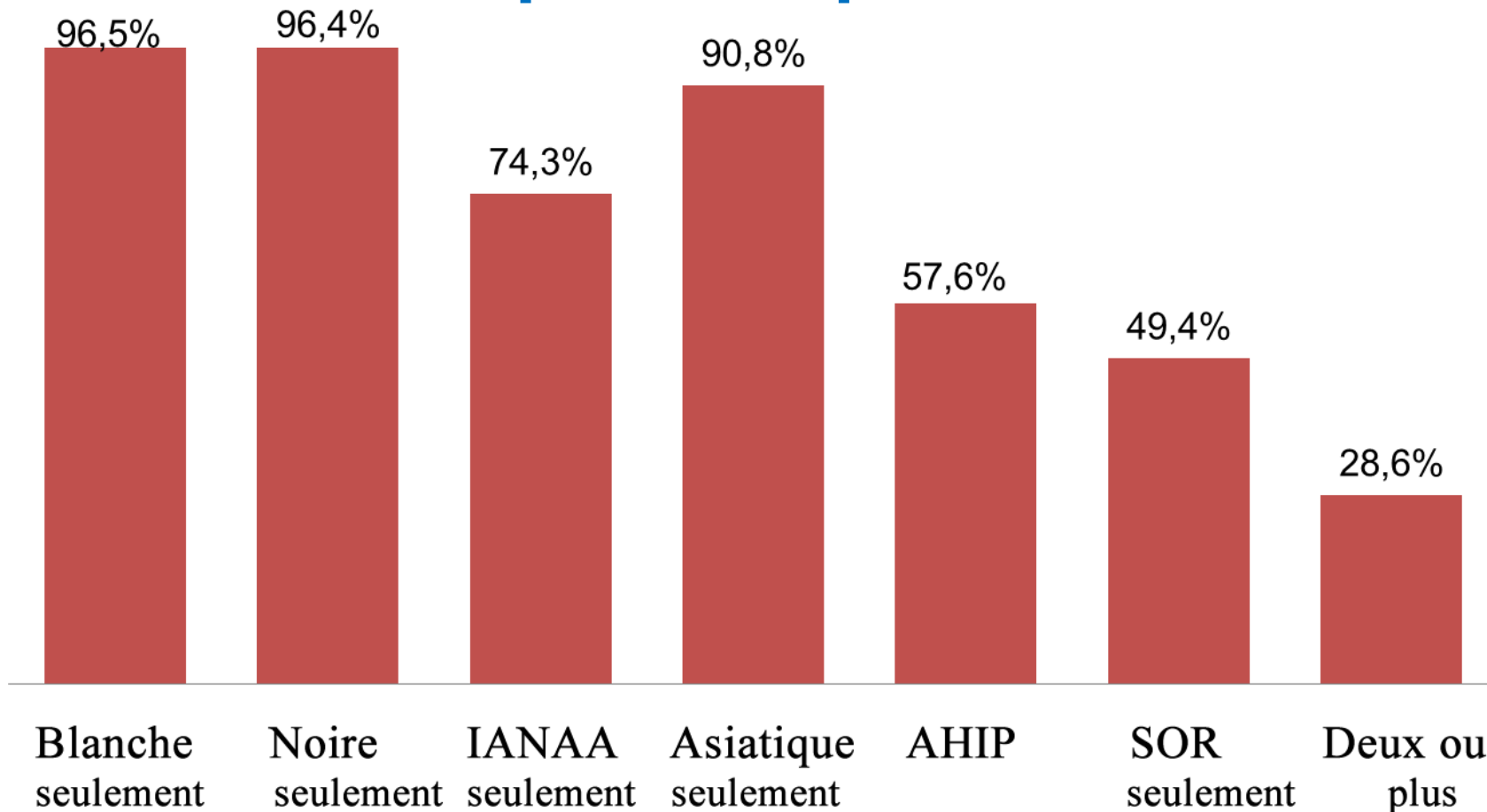
Autres méthodes utilisées pour attribuer la race dans les DA

- Race dans les DA, méthode 5
 - Race attribuée à partir des dossiers du recensement précédent. Si aucune réponse dans les dossiers du recensement précédent, la race a été attribuée selon la méthode 1.
- Race dans les DA, méthode 6
 - Si une réponse IANAA est présente dans les données de l'IHS, cette réponse a été attribuée à l'enregistrement du DA. Autrement, une réponse pour la race a été attribuée selon la méthode 5.

Pourcentage de données couplées du recensement et des DA qui correspondent pour l'origine hispanique



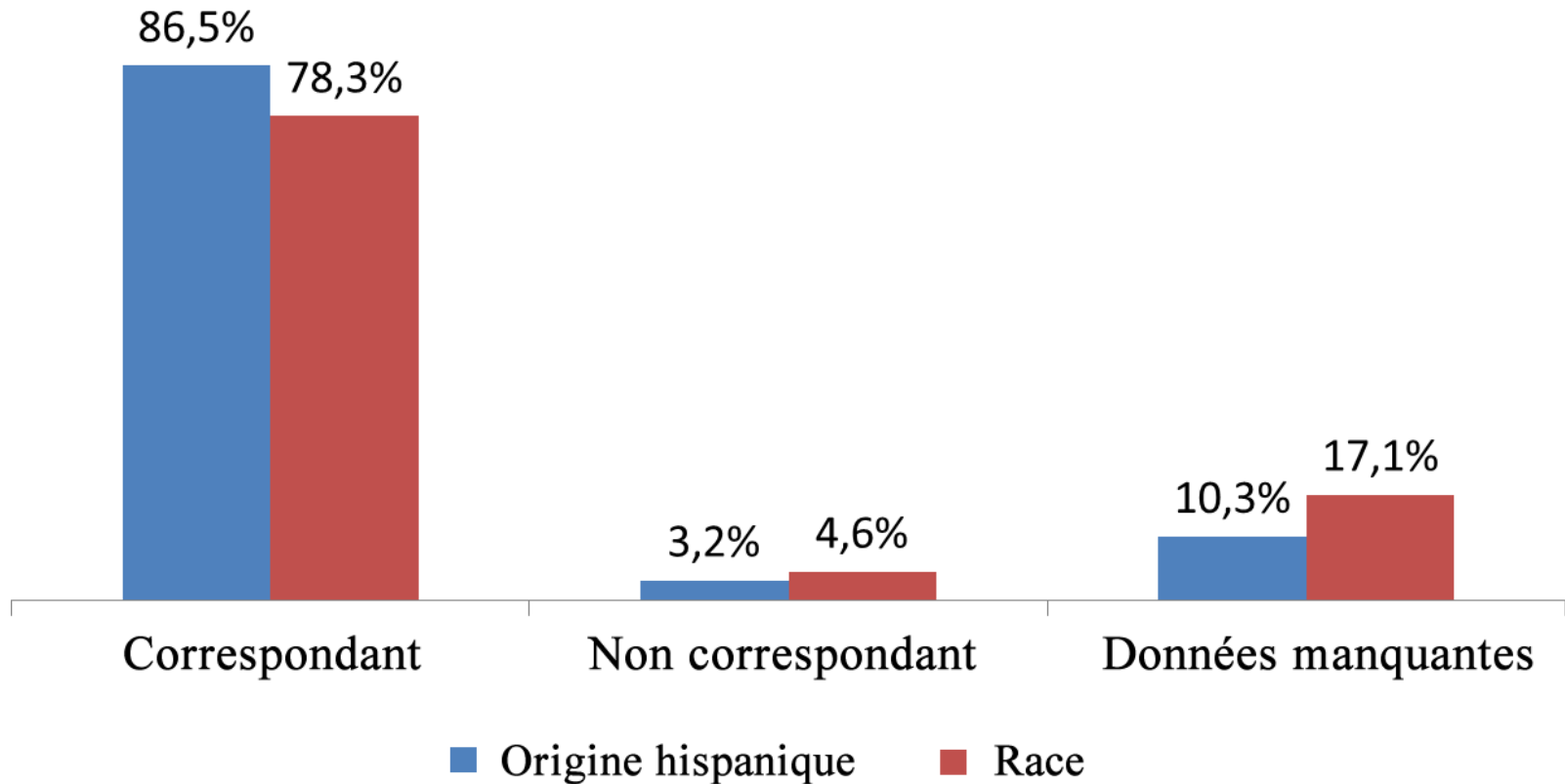
Pourcentage de données couplées du recensement et des DA qui correspondent pour la race



Méthodologie

- Régression logistique multinomiale
 - La variable dépendante comporte trois catégories
 - Réponse correspondant (référence)
 - Réponse ne correspondant pas
 - Réponse manquante dans les DA
 - Régressions distinctes pour l'origine hispanique et la race

Distribution des variables dépendantes, données couplées du recensement et des DA



Source : Recensement de 2010 et dossiers administratifs

Certains résultats de la régression logistique multinomiale, rapports de cotes

	Modèle pour l'origine hispanique		Modèle pour la race	
	Réponses ne correspondant pas	Manquantes dans les DA	Réponses ne correspondant pas	Manquantes dans les DA
Origine ethnique/Race dans le recensement (Blanche NH seulement mise de côté)				
Hispanique	1,58***	0,71***	42,64***	11,24***
Noire NH seulement	0,67***	0,88***	1,63***	0,71***
IANAA NH seulement	2,20***	0,89***	13,67***	0,41***
Asiatique NH seulement	2,51***	2,98***	6,29***	10,88***
AHIP NH seulement	3,60***	2,01***	35,95***	9,24***
Âge (18 à 44 ans laissés de côté)				
0 à 17 ans	1,24***	12,66***	2,19***	16,85***
45 à 64 ans	0,96***	0,40***	0,95***	0,48***
65 ans et plus	0,83***	0,24***	0,86***	0,25***

Modèle pour l'origine hispanique n : 251 320 952

Modèle pour la race n : 253 905 696

NH = Non hispanique ***p<,001

Certains résultats de la régression logistique multinomiale, rapports de cotes

	Modèle pour l'origine hispanique		Modèle pour la race	
	Réponses ne correspondant pas	Manquantes dans les DA	Réponses ne correspondant pas	Manquantes dans les DA
Origine ethnique/Race dans le recensement (Blanche NH seulement mise de côté)				
Hispanique	1,58***	0,71***	42,64***	11,24***
Noire NH seulement	0,67***	0,88***	1,63***	0,71***
IANAA NH seulement	2,20***	0,89***	13,67***	0,41***
Asiatique NH seulement	2,51***	2,98***	6,29***	10,88***
AHIP NH seulement	3,60***	2,01***	35,95***	9,24***
Âge (18 à 44 ans laissés de côté)				
0 à 17 ans	1,24***	12,66***	2,19***	16,85***
45 à 64 ans	0,96***	0,40***	0,95***	0,48***
65 ans et plus	0,83***	0,24***	0,86***	0,25***

Modèle pour l'origine hispanique n : 251 320 952

Modèle pour la race n : 253 905 696

NH = Non hispanique ***p<,001

Certains résultats de la régression logistique multinomiale, rapports de cotes

	Modèle pour l'origine hispanique		Modèle pour la race	
	Réponses ne correspondant pas	Manquantes dans les DA	Réponses ne correspondant pas	Manquantes dans les DA
Origine ethnique/Race dans le recensement (Blanche NH seulement mise de côté)				
Hispanique	1,58***	0,71***	42,64***	11,24***
Noire NH seulement	0,67***	0,88***	1,63***	0,71***
IANAA NH seulement	2,20***	0,89***	13,67***	0,41***
Asiatique NH seulement	2,51***	2,98***	6,29***	10,88***
AHIP NH seulement	3,60***	2,01***	35,95***	9,24***
Âge (18 à 44 ans laissés de côté)				
0 à 17 ans	1,24***	12,66***	2,19***	16,85***
45 à 64 ans	0,96***	0,40***	0,95***	0,48***
65 ans et plus	0,83***	0,24***	0,86***	0,25***

Modèle pour l'origine hispanique n : 251 320 952

Modèle pour la race n : 253 905 696

NH = Non hispanique ***p<,001

Certains résultats de la régression logistique multinomiale, rapports de cotes

	Modèle pour l'origine hispanique		Modèle pour la race	
	Réponses ne correspondant pas	Manquantes dans les DA	Réponses ne correspondant pas	Manquantes dans les DA
Origine ethnique/Race dans le recensement (Blanche NH seulement mise de côté)				
Hispanique	1,58***	0,71***	42,64***	11,24***
Noire NH seulement	0,67***	0,88***	1,63***	0,71***
IANAA NH seulement	2,20***	0,89***	13,67***	0,41***
Asiatique NH seulement	2,51***	2,98***	6,29***	10,88***
AHIP NH seulement	3,60***	2,01***	35,95***	9,24***
Âge (18 à 44 ans laissés de côté)				
0 à 17 ans	1,24***	12,66***	2,19***	16,85***
45 à 64 ans	0,96***	0,40***	0,95***	0,48***
65 ans et plus	0,83***	0,24***	0,86***	0,25***

Modèle pour l'origine hispanique n : 251 320 952

Modèle pour la race n : 253 905 696

NH = Non hispanique ***p<,001

Certains résultats de la régression logistique multinomiale, rapports de cotes

	Modèle pour l'origine hispanique		Modèle pour la race	
	Réponses ne correspondant pas	Manquantes dans les DA	Réponses ne correspondant pas	Manquantes dans les DA
Origine ethnique/Race dans le recensement (Blanche NH seulement mise de côté)				
Hispanique	1,58***	0,71***	42,64***	11,24***
Noire NH seulement	0,67***	0,88***	1,63***	0,71***
IANAA NH seulement	2,20***	0,89***	13,67***	0,41***
Asiatique NH seulement	2,51***	2,98***	6,29***	10,88***
AHIP NH seulement	3,60***	2,01***	35,95***	9,24***
Âge (18 à 44 ans laissés de côté)				
0 à 17 ans	1,24***	12,66***	2,19***	16,85***
45 à 64 ans	0,96***	0,40***	0,95***	0,48***
65 ans et plus	0,83***	0,24***	0,86***	0,25***

Modèle pour l'origine hispanique n : 251 320 952

Modèle pour la race n : 253 905 696

NH = Non hispanique ***p<,001

Conclusion et prochaines étapes

- Constatations similaires par rapport aux travaux de 2000, réponses dans les DA pour les groupes minoritaires plus faibles par rapport aux Blancs non hispaniques
- Utiliser d'autres données fédérales, d'état, et de tiers pour combler les lacunes lorsque la correspondance entre la race et l'origine hispanique dans les DA est faible par rapport au Recensement de 2010
- Règles opérationnelles additionnelles pour la recherche
- Distinguer les problèmes de fluidité de l'origine hispanique et de la race des problèmes de qualité des données

Merci!

sharon.r.ennis@census.gov

sonya.rastogi@census.gov

james.noon@census.gov