ترميز الإجابات المكتوبة في التعدادات السكانية

مواضيع مختارة في التعدادات السكانية الدولية[1]

صدر في يونيو 2020

مقدمة

تشتمل جميع الأدوات المستخدمة في إجراء التعدادات السكانية على أسئلة تتطلب من المشاركين تقديم إجابات مكتوبة. تعطي الحقول الفارغة المخصصة لتدوين الإجابات المكتوبة الفرصة للمشاركين للإجابة من خلال كتابة نص بأسلوب حر بدلاً من الاختيار من بين فئات الإجابات المحددة مسبقًا. تشجِّع الإجابات المكتوبة المشاركين من خلال السماح لهم بالإجابة عن الأسئلة بحرية وعدم التقيد بفئات الإجابات المُعدة مسبقًا. لكن إدراج حقول الإجابات المكتوبة في الاستبيان تنتج عنه مجموعة معينة من التحديات التي تواجه مكاتب الإحصاء الوطنية (NSO) أثناء جمع البيانات ومعالجتها.

تُمثل هذه الوثيقة الفنية الإرشادية إحدى حلقات سلسلة "مواضيع مختارة في التعدادات السكانية الدولية (STIC) "، وتقدم معلومات إلى مكاتب الإحصاء الوطنية (NSO) عن المعايير المعترف بها دوليًا للترميز الآلي للإجابات المكتوبة في التعدادات السكانية.

صيغ الأسئلة

تُعد الأسئلة المُقيَّدة (مُحدَّدة الإجابة) هي النوع الأكثر شيوعًا في استبيانات التعداد السكاني. وتتضمن هذه الأسئلة قائمة ببدائل الإجابات ذات الخيارات الثابتة والمحددة، وعادةً ما تطلب من المشاركين تحديد خيار أو أكثر من تلك الخيارات. في المقابل، لا تحتوي الأسئلة الحرة (مفتوحة الإجابة) على فئات للإجابات المحددة مسبقًا، وتتيح للمشاركين الإجابة عن الأسئلة بحرية كلٌّ حسب أسلوب تعبيره. وفي استبيانات التعداد السكاني الورقية والتي تكون مصممة للمسح الضوئي، تُكتَب الإجابات الحرة عادةً في حقول "الإجابات المكتوبة"، التي يحدد حجمها الافتراضي الحد الأقصى لعدد الأحرف المسموح بها في الإجابة.

لكلٍّ من الأسئلة المُقيَّدة والأسئلة الحرة مزايا وعيوب. عادةً ما تكون الأسئلة المُقيَّدة التي تتضمن قائمة بالإجابات الثابتة أسهل في الإجابة والتحليل؛ ما يقلل العبء الواقع على المشاركين ومكاتب الإحصاء الوطنية (NSO) أثناء جمع البيانات ومعالجتها وتحليلها. وبالرغم من ذلك، غالبًا ما تفتقر فئات الإجابات المحددة مسبقًا إلى مجموعة واسعة من الإجابات التي كان من الممكن أن يقدمها المشاركون لو طُرِح السؤال بصيغة حرة ومفتوحة. في المقابل، تتيح الأسئلة الحرة أو المكتوبة للمشاركين التعبير عن أنفسهم بأسلوبهم بناءً على تمام معرفتهم وفهمهم. لكن لهذا النوع من الأسئلة عيوبه أيضًا، إذ يصعب على مكاتب الإحصاء الوطنية (NSO) معالجة الإجابات عن الأسئلة الحرة وتحليلها. قد تكون الأسئلة الحرة أيضًا مصدرًا قويًا للتحيز، لأنه حتى إذا صيغ السؤال صياغةً صحيحة، فقد تختلف طريقة فهمه باختلاف المشاركين.

عادةً ما يطرح الباحثون في مكاتب الإحصاء الوطنية (NSO) أسئلةً حرةً عندما تكون المتاعب المصاحبة للأسئلة الحرة أقل من القيود الخاصة بالأسئلة المُقيَّدة. وفي حالات أخرى، يجب طرح الأسئلة بصيغة حرة، كما هو موضح أدناه.

[1] تُمثل هذه الوثيقة الإرشادية إحدى حلقات سلسلة *"مواضيع مختارة في التعدادات السكانية الدولية"*، والتي تتناول بعمق المسائل التي تثير اهتمام المجتمع الإحصائي على الصعيد الدولي. ويقوم مكتب الإحصاء الأمريكي بدور فاعل في دعم الدول لتطوير أنظمتها الإحصائية الوطنية، مستثمرًا في تنمية القدرات والارتقاء بالخبرات الإحصائية على نحو يحقق الاستدامة والتطور المستمرين.

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

United States Agency for International Development
www.usaid.gov

United Nations Population Fund
www.unfpa.org

الجدول رقم 1.

النشاط الاقتصادي والمهنة ومجال العمل: تعداد السكان والمساكن في بوتسوانا عام 2011

جميع الأشخاص البالغين من العمر 12 عامًا فأكبر					
النشاط الاقتصادي				المهنة	مجال العمل
ما العمل الذي كان... يؤديه بشكل أساسي منذ عيد الاستقلال عام 2010؟ **عمل موسمي** 01 مدفوع الأجر 02 غير مدفوع الأجر **عمل غير موسمي** 03 مدفوع الأجر 04 غير مدفوع الأجر **أخرى** 05 باحث عن عمل 06 عمل منزلي 07 طالب 08 متقاعد 09 مريض غير ذلك (يُرجى التوضيح)	هل أدى... أي نوع من العمل مقابل أجر أو ربح أو فائدة أو استهلاك منزلي لمدة ساعة واحدة على الأقل في الأيام السبعة الماضية؟ 1 نعم (انتقل إلى A22) 2 لا [إذا كانت الإجابة بـ "لا"، فهل كان...يعمل في زراعة أرضه/يرعى ماشيته؟]	بما أن... لم يكن يمارس أي عمل، فماذا كان يفعل؟ 1 البحث عن عمل بنشاط 2 عمل منزلي 3 طالب 4 متقاعد 5 مريض غير ذلك (يُرجى التوضيح) **[إذا كانت أنثى فانتقل إلى A25** **إذا كان ذكرًا فانتقل إلى الشخص التالي]**	ماذا كان يعمل... خلال الأيام السبعة الماضية؟ 1 موظف - بأجر نقدي 2 موظف - بأجر عيني 3 صاحب عمل حر (ليس لديه موظفون) 4 صاحب عمل حر (لديه موظفون) 5 فرد من العائلة يساعد عائلته في أعمالها من دون أجر 6 العمل في زراعة أرضه/ رعاية ماشيته	ما نوع العمل الذي كان... يؤديه في الأيام السبعة الماضية؟ على وجه الدقة، ما المهام والواجبات الرئيسية التي كان يؤديها؟ *(ينبغي التدقيق عند الضرورة، يُرجى استخدام كلمتين أو أكثر لوصف المهنة)*	ما المنتج أو الخدمة أو النشاط الرئيسي لمكان العمل الذي كان يعمل فيه...؟ *(ينبغي التدقيق عند الضرورة، يُرجى استخدام كلمتين أو أكثر لوصف مجال العمل)* **انتقل إلى A25 للإناث، عدا ذلك** انتقل إلى الشخص التالي
A19(2)	A20	A21	A22	A23(3)	A24(4)

المصدر: Statistics Botswana, 2011.

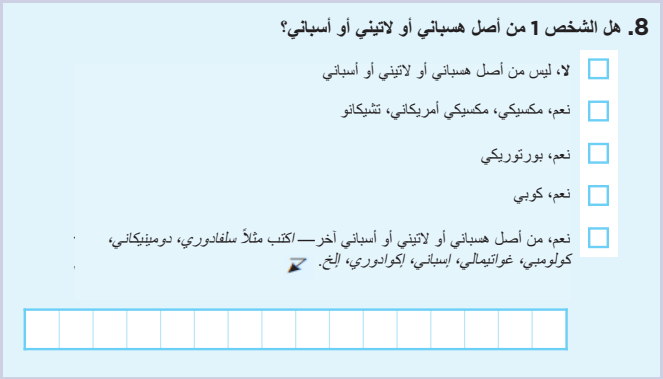
الأسئلة ذات الصيغ الحرة والأسئلة ذات الصيغ المختلطة

تتضمن جميع استبيانات التعداد السكاني سؤالاً واحدًا مكتوبًا على الأقل- وهو السؤال عن الاسم الأول واسم العائلة فيما يتعلق بسكان البيت. بالإضافة إلى ذلك، تسمح مكاتب الإحصاء الوطنية (NSO) في الغالب بتقديم إجابات مكتوبة على بعض أسئلتها. على سبيل المثال، غالبًا ما تكون الإجابات عن الأسئلة المتعلقة بمجال العمل والمهنة حرة. من الصعب ترميز أو إعداد قائمة بمئات المهن المتضمنة في قائمة رموز مجالات العمل والمهن، لا سيما في حالة الاستبيانات الورقية. ومن ثم، فغالبًا ما تسمح مكاتب الإحصاء الوطنية (NSO) بتقديم إجابات مكتوبة عندما تريد الحصول على تفاصيل أكثر من الحد العملي لعرض التفاصيل في الاستبيانات الورقية أو الرقمية. يقدم الجدول رقم 1 مثالاً على قسم العمل في أحد التعدادات السكانية (مُقتبَس من التعداد السكاني في بوتسوانا لسنة 2011). يُرجى ملاحظة أن أسئلة النشاط الاقتصادي تُطرَح بصيغة مُقيَّدة، بينما تُطرَح الأسئلة المتعلقة بالمهنة والمجال بصيغة حرة.

يوضح الشكل رقم 1 مثالاً على سؤال من التعداد السكاني للولايات المتحدة لسنة 2020، حيث تم دمج أربعة مربعات اختيار وحقل إجابة مكتوبة لجمع معلومات مفصلة عن الأشخاص

الشكل رقم 1.

سؤال ذو صيغة مختلطة: نموذج البيوت من التعداد السكاني للولايات المتحدة لسنة 2020



المصدر: U.S. Census Bureau, 2020 Census.

ذوي الأصل الهسباني أو اللاتيني. يُخصَّص مربع الاختيار الأول للمشاركين الذين يختارون الإجابة "ليس من أصل هسباني أو لاتيني أو أسباني". وتُخصَّص مربعات الاختيار الثلاثة التالية لأكبر المجموعات ذات الأصل الهسباني - "مكسيكي" و"بورتوريكي" و"كوبي". أما مربع الاختيار الأخير: "من أصل هسباني أو لاتيني أو أسباني آخر"، فيكون مصحوبًا بمجموعة من الأمثلة المفصَّلة وحقل مُخصَّص لتقديم إجابة مكتوبة لجمع البيانات المتعلقة بجميع المجموعات الأخرى من السكان ذوي الأصول الهسبانية أو اللاتينية. يتيح الجمع بين مربعات الاختيار المُفصَّلة وحقل الإجابة المكتوبة لجميع المشاركين فرصة الإبلاغ عن أصلهم الهسباني. يُعَد هذا التصميم حلاً وسطًا بين الأسئلة المُقيَّدة والأسئلة الحرة، حيث تعتمد هذه الصيغة على وجود سؤال حر بصيغة سؤال مُقيَّد.

تختبر مكاتب الإحصاء الوطنية (NSO) في بعض الأحيان صيغًا مختلفة للأسئلة، لتجد الصيغة المثلى التي تساعدها في جمع البيانات عالية الجودة التي تريدها، وفي الوقت المحدد، ومن دون استنزاف ميزانياتها. غالبًا ما تؤتي صيغ الأسئلة المبتكرة ثمارها عند تنفيذها بشكل صحيح.

حقول "الاسم" المكتوبة

يعتبر السؤال عن "الاسم" فريدًا من حيث إنه لا يمكن طرحه إلا من خلال صيغة سؤال حر، وليس المقصود منه أن يكون وسيلة للحصول على معلومات ذات صلة بالتعداد السكاني أو الاستبيان. إنما يُستخدَم في المقام الأول باعتباره مُحدِّدًا للهوية. غالبًا ما تُستخدم الأسماء لتحديد هوية المشاركين أثناء جمع البيانات من بيت أو أسرة ما، ولأغراض المطابقة في استبيانات ما بعد التعداد (PES). تُستخدم الأسماء أيضًا لأغراض الفهرسة في أدوات استرجاع البيانات من محفوظات التعدادات السكانية (راجع "مواضيع مختارة في التعدادات السكانية الدولية (STIC)": موضوع "أرشفة بيانات التعداد السكاني وحفظها").

هناك سمة أخرى تميز إجابات "الاسم" المكتوبة في التعداد السكاني، وهي عدم الحاجة إلى ترميزها. ومع ذلك، يجب معالجة الاختلافات الإملائية التي تنتج أثناء جمع البيانات أو استخلاصها حتى تُستخدَم الأسماء بشكل صحيح كمحددات للهوية، إذ غالبًا ما تُستخدَم الأسماء لأغراض المطابقة في استبيانات ما بعد التعداد. على سبيل المثال، قد يُخطئ المحاورون في الإملاء الصحيح للاسم أثناء إجراء المقابلات الشخصية أو الهاتفية. يشيع حدوث ذلك مع الأسماء التي يكون لها أكثر من هجاء، أو مع الأسماء التي تتشابه في النطق الصوتي وتختلف في الهجاء. على سبيل المثال، تمثل الكلمات Sean وShaun وShawn وShon طرق هجاء مختلفة لاسم واحد؛ كما تُنطق أسماء العائلات Meier وMeyer وMaier وMayer وMair وMayr بنفس الطريقة.

بمجرد أن تصبح سجلات التعداد السكاني قابلة للقراءة آليًا، يجب فهرسة الأسماء لتسهيل ربط السجلات الشخصية واسترجاعها. يعرض المربع رقم 1 مقدمة عن نظام Soundex، وهو من أقدم أنظمة الفهرسة الصوتية وأكثرها شهرة.

المربع رقم 1.

الفهرسة الصوتية ونظام Soundex

غالبًا ما يُعد نظام Soundex، الذي وضعه راسل (Russell) وأوديل (Odell) في عام 1918، أول نظام للفهرسة الصوتية لأسماء العائلات. ومع حلول ثلاثينيات القرن العشرين، أصبح نظام Soundex مستخدمًا بالفعل لفهرسة بيانات التعداد السكاني في الولايات المتحدة. ولا تزال "إدارة المحفوظات والسجلات الوطنية" (NARA) الأمريكية تستخدم هذا النظام للبحث عن السجلات في محفوظات التعداد السكاني السابقة. يتميز نظام Soundex بأنه مرن إلى درجة تسمح بمواءمة مبادئه الأساسية مع اللغات المختلفة واستخدامات مختلفة. فيما يلي نسخة أسبانية من نظام Soundex.

يتكون الرمز الأصلي في نظام Soundex من حرف واحد وثلاثة أرقام. الحرف هو الحرف الأول من اسم العائلة، وتُخصَّص الأرقام وفقًا للجدول الموضح أدناه (إذا لم يكن اسم العائلة طويلًا بما يكفي لإنتاج ثلاثة أرقام، تُضاف أصفار متتابعة). ويتم تجاهل حروف العلة (الحروف المتحركة) في هذه الخوارزمية، ويتم تجميع الحروف الساكنة مع الحروف الساكنة المشابهة لها بناءً على مخارجها الصوتية. تنطبق بعض القواعد الإضافية على الأسماء التي تحتوي على أحرف مزدوجة، والبادئات، والأحرف المتجاورة التي لها نفس الرقم في نظام Soundex، والأسماء المتعددة الكلمات، وغيرها. يُرجى ملاحظة أن أسماء العائلات Meier وMeyer وMaier وMayer وMair وMayr (المذكورة أعلاه) تُنتِج جميعها نفس الرمز، M600، حسب نظام Soundex.

الأصل (اللغة الإنجليزية)

نظام الترميز SOUNDEX

الرمز	الحرف
1	B, F, P, V
2	C, G, J, K, Q, S, X, Z
3	D, T
4	L
5	M, N
6	R
يُرجى تجاهل حروف:	**A, E, H, I, O ,U, W, Y**

المصدر: Russell, 1918.

يتوافق الجدول التالي مع خوارزمية "PhoneticSpanish"، وهي عبارة عن تعديل لنظام Soundex ليناسب اللغة الأسبانية. يستخدم هذا الجدول الأرقام من 0 إلى 9 بدلاً من الأرقام من 1 إلى 6؛ ولا يتم تجاهل الحرفين "Y" و"H"؛ وتتم إضافة الحروف الأسبانية "LL" و"Ñ" و"RR" إلى الخوارزمية. وتُطبَّق قواعد خاصة على بعض مجموعات الحروف. يتكون الرمز في خوارزمية "PhoneticSpanish" من أرقام فقط، ولا يقتصر على أربعة أحرف.

الترميز الصوتي الإسباني

الرمز	الحرف
0	P
1	B, V
2	F, H
3	D, T
4	C, S, X, Z
5	L, LL, Y
6	M, N, Ñ
7	K, Q
8	G, J
9	R, RR
يُرجى تجاهل حروف	**A, E, I, O, U, W**

المصدر: Amón, Moreno & Echeverri, 2012.

أنظمة الترميز

بمجرد أن تتم رقمنة الإجابات المكتوبة، يجب تصنيفها إلى فئات مُرقَّمة محددة مسبقًا، أو ترميزها. ويجب أيضًا ترميز الإجابات المكتوبة التي يتم جمعها بالوسائل الرقمية. هناك ثلاث طرق رئيسية للترميز: الترميز الكتابي، والترميز بمساعدة الحاسوب، والترميز الآلي.

الترميز الكتابي

يُعد الترميز الكتابي أو اليدوي النظام الأقل تطوراً بين أنظمة الترميز. الترميز الكتابي هو "الترميز اليدوي" ويعتمد تمامًا على القائمين على الترميز المكلفين بإتمام المهمة بمساعدة الأدلة وقوائم الرموز فقط. ويمكن إجراء ذلك النوع من الترميز من خلال المحاورين في الميدان أثناء جمع البيانات، أو من خلال القائمين على الترميز المدربين أثناء معالجة البيانات. وفي بعض الحالات، يقوم المشاركون أنفسهم بإجراء الترميز الكتابي عند إرفاق قوائم الرموز بالاستبيان.

لا يعد الترميز الكتابي للإجابات المكتوبة في التعدادات السكانية خيارًا عمليًا، نظرًا إلى كمية الإجابات الكبيرة. وغالبًا ما تكون أنظمة الترميز الكتابي مُكلِّفة، وتستغرق وقتًا طويلاً، حتى في أصغر البلدان.

الترميز بمساعدة الحاسوب

بالإضافة إلى الترميز الكتابي، يمكن ترميز الإجابات المكتوبة باستخدام نظام الترميز بمساعدة الحاسوب (CAC). يتعاون القائمون على الترميز البشريون في نظام الترميز بمساعدة الحاسوب (CAC) بصورة تفاعلية مع جهاز الحاسوب أثناء تعيين الرموز. يشبه نظام الترميز بمساعدة الحاسوب (CAC) نظام الترميز الكتابي. لكن القائم على الترميز في نظام الترميز بمساعدة الحاسوب (CAC) يتمتع بإمكانية الوصول إلى عدد من موارد الحوسبة مثل شاشات المساعدة وجداول القرارات والمعلومات الإضافية، وغير ذلك من الموارد.

تعد أنظمة الترميز بمساعدة الحاسوب (CAC) أكثر تقدمًا من أنظمة الترميز الكتابي، وغالبًا ما توفر الوقت والمال لمكاتب الإحصاء الوطنية (NSO). كما تؤدي أنظمة الترميز بمساعدة الحاسوب (CAC) دورًا مهمًا في تقييم أنظمة الترميز الآلي وتحسينها، وذلك على النحو الموضح أدناه.

الترميز الآلي

تعد أنظمة الترميز الآلي الأنظمة الأكثر تقدمًا من الناحية التكنولوجية بين أنواع أنظمة الترميز الثلاثة. يشغِّل المبرمجون في أنظمة الترميز الآلي برنامجًا بمواصفات مُفصَّلة تُستخدم لتصنيف الإجابات (أي ترميزها). ويعالج الإجابات المتبقية، التي يتعذر على النظام ترميزها، الأشخاص القائمون على الترميز في أنظمة الترميز بمساعدة الحاسوب (CAC). تُستخدَم بعد ذلك الإجابات المُعالَجة في نظام الترميز بمساعدة الحاسوب بشكل تكراري لتحسين قدرة النظام على تصنيف المزيد من الإجابات تلقائيًا.

تقلِّل برامج الترميز الآلية تكاليف المعالجة، وتضمن تطبيق قواعد الترميز على نحوٍ منتظم. ورغم ذلك، تتطلب الأنظمة الآلية، بل وحتى أكثرها تطورًا، قدرًا معينًا من الترميز الكتابي. وتعد أنظمة الترميز الكتابي أو أنظمة الترميز بمساعدة الحاسوب ضرورية في المراحل الأولى من تطوير برامج الترميز الآلي، لضمان عملها بشكل صحيح واستمرار تحديثها.

أنظمة الترميز الحديثة

تتطلب التعدادات السكانية الحديثة الجمع بين استخدام الترميز الآلي والكتابي لضمان إنتاج بيانات عالية الجودة على نحوٍ أقل تكلفة.

وتتضمن عملية بناء نظام الترميز الآلي خمسة أجزاء أو خمس مراحل أساسية (انظر الشكل رقم 2).

المرحلة 1: بناء القاموس

تشبه قواميس أنظمة الترميز التلقائي قوائم الرموز، باستثناء أن القواميس تتضمن أكبر عدد ممكن من الأشكال المختلفة لنفس الرمز. قد يحتوي قاموس نظام الترميز الآلي المُستخدَم في التعداد السكاني على مئات الآلاف من المداخل، حسب حجم البلد الذي يُجري التعداد السكاني. يجب أن يتناول القاموس الاختلافات الشائعة في الهجاء واللغة، وأن يتضمن الاختلافات الناتجة عن الأخطاء المطبعية أو أخطاء المسح الضوئي التي تنتج عن الكتابة اليدوية الرديئة أو غير الواضحة. على سبيل المثال، قد يتضمن سؤال افتراضي في القاموس عن الأصل بعض المداخل مثل: Affrican، وAfrcn، وAfricaine، وAfricam، وAfricano، وAfricsn، وAfrika، وi-african، وSfrican، وما إلى ذلك. تمثل جميع هذه المداخل أخطاء مطبعية، وأخطاء إملائية، وكلمات أجنبية مكافئة لمفردة "African" (أفريقي)، ويجب ترميزها بنفس الطريقة. يجب بناء القواميس بمرور الوقت، إذ إنها نادرًا ما تستخلص أي قدر ذي معنى من الإجابات المحتملة في البداية.

المرحلة 2: إدخال بيانات الإجابات

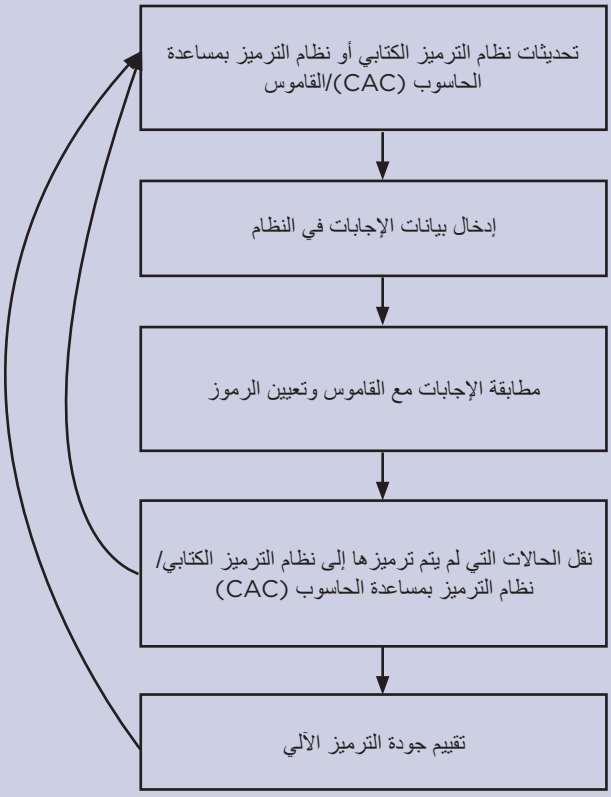
تُدخَل البيانات المكتوبة المرقمنة في النظام دون تصحيحات أو تعديلات. تتمثل التعديلات الوحيدة المسموح بإجرائها على الإجابات المكتوبة في هذه المرحلة في إزالة المفردات والأحرف غير الصالحة.

تتضمن المفردات غير الصالحة كلمات أو مجموعات من الكلمات لا تشكل إجابات صحيحة، ولا تغير معنى الإجابة عند إزالتها. على سبيل المثال، في سؤال افتراضي عن الأصل العرقي، يجب ترميز الإجابة "نصف أفريقي ونصف أوروبي" على أنها "أفريقي" و"أوروبي" على حدة. يمكن إزالة كلمة "نصف" وحرف العطف "و" من دون تغيير معنى الإجابة المقصودة. لكن في الإجابة "ليس أفريقيًا، أوروبيًا"، على الرغم من أن مفردة "ليس" لا تُعد إجابة صحيحة فيما يتعلق بالأصل العرقي، إلا أنه لا يمكن حذفها من دون تغيير معنى الإجابة المقصودة. يجب عدم تجاهل مفردة "ليس"، بل استخدامها بدلاً من برمجة قاعدة تعديل لتجاهل المفردة المجاورة لها أيضًا. ولذا، يجب، في هذا المثال، ترميز كلمة "أوروبي" فقط كإجابة صحيحة.

تشمل الأحرف غير الصالحة الفواصل، والفواصل العليا، وعلامات الاستفهام أو التعجب، والرموز الرياضية، وجميع الأحرف الأخرى التي يعتبرها مكتب الإحصاء الوطني غير صالحة. تنطبق إزالة الأحرف غير الصالحة فقط على الإجابات المستخلصة من الاستبيانات الورقية المرقمنة، حيث يجب اعتماد الأحرف الصالحة فقط في أشكال التجميع الإلكترونية مثل التطبيقات القائمة على الويب أو تطبيقات الأجهزة اللوحية. تتم، في كثير من الأحيان، إزالة الأحرف غير الصالحة أثناء المسح الضوئي أو عملية الرقمنة.

الشكل رقم 2.

نظام الترميز الآلي



المصدر: U.S. Census Bureau.

المرحلة 3: مطابقة الإجابات

بمجرد إزالة المفردات والأحرف غير الصالحة من مداخل الإجابات، تحاول أنظمة الترميز الآلي مطابقة ما تبقى من السلاسل النصية بالرموز المخصصة من مكتب الإحصاء الوطني باستخدام القاموس. من الناحية النظرية، يجب أن يكون لجميع الإجابات ما يطابقها في القاموس. لكن من الناحية العملية يبدو ذلك مستحيلًا. في مرحلة مطابقة الإجابات، تُعطى الإجابات، التي توجد مداخل مطابقة لها في القاموس، رمزًا مخصصًا، وتوضع علامة على الإجابات التي لا يستطيع النظام مطابقتها، على أنها إجابات متبقية.

المرحلة 4: مطابقة الحالات المتبقية

يتم نقل الإجابات المتبقية إلى أنظمة الترميز بمساعدة الحاسوب لكي يعالجها المُرمِّزون المدرَّبون، الذين غالبًا ما يتخصصون في سؤال أو أكثر من الأسئلة المحددة. يجب الأخذ في الحسبان أنه من المتوقع أن تكون الحالات المتبقية أكثر صعوبة في الترميز وأقل شيوعًا من الحالات العادية. لهذا السبب، من الشائع أن يتبع القائمون على الترميز إجراءات محددة لإحالة الحالات الأكثر صعوبة إلى خبراء متخصصين من أجل الترميز أو التحكيم. وبمجرد ترميز الإجابات المتبقية، يُحدَّث القاموس بإضافة النظير الجديد المطابق لرمز الإجابة المخصص من مكتب الإحصاء الوطني. وبهذه الطريقة، سيتمكن النظام من مطابقة هذه الإجابات وترميزها تلقائيًا في المستقبل.

المرحلة 5: تقييم الجودة

تساعد مرحلة تقييم الجودة مكاتب الإحصاء الوطنية (NSO) على تقييم جودة الأنظمة الآلية من أجل تحسينها. ولتحقيق ذلك، تخضع عينة من الإجابات التي نجح ترميزها آليًا للترميز اليدوي على يد كبار القائمين على الترميز. ومن ثم يمكن للمبرمجين إضافة قواعد ترميز جديدة، أو إجراء تصحيحات على النظام الآلي، أو تعديل القاموس إذا لزم الأمر.

يلخص المربع رقم 2 عملية ترميز التعداد السكاني التجريبي لسنة 2019 التي أجراها مكتب الإحصاء الأمريكي للإجابات المكتوبة المتعلقة بالأصل والعرق الهسباني.

المربع رقم 2.

عملية الترميز وإجراءات الإحالة الخاصة بمكتب الإحصاء الأمريكي

صمَّم مكتب الإحصاء الأمريكي العديد من أنظمة الترميز الآلي للتعدادات السكانية المختلفة، والاستبيانات لمختلف الأسئلة. ومع ذلك، تبقى المبادئ الأساسية ثابتة. فيما يلي موجز لإجراءات الترميز الخاصة بالأسئلة المتعلقة بالأصل والعرق الهسباني في التعداد السكاني التجريبي لسنة 2019.

- في التعداد السكاني التجريبي لعام 2019، تألفت عملية الترميز من مرحلتين رئيسيتين: (1) الترميز الآلي و(2) ترميز الإجابات المتبقية.
- أثناء عملية الترميز الآلي، تم ترميز الإجابات المكتوبة المُجمَّعة حديثًا من خلال مقارنتها بالسجلات الموجودة في القاموس أو "الملف الرئيسي"، الذي يحتوي على مئات الآلاف من مداخل الإجابات المكتوبة المُرمَّزة سابقًا، والتي تراكمت على مدار العديد من التعدادات والاستبيانات. تم تعيين الرمز متى وُجِد التطابق. تطلبت الإجابات المكتوبة، التي لم يوجد لها رمز يطابقها، تخصيص مسؤول ترميز مدرب لتعيين الرموز المناسبة يدويًا.
- تضمنت عملية ترميز حالات الإجابات المتبقية ثلاثة أنشطة رئيسية:

أ. صياغة الرموز - تم ترميز الإجابات المتبقية يدويًا.

ب. التحقق من الرموز - تم ترميز عينة من الحالات مرة ثانية بواسطة مسؤول ترميز مختلف.

ج. التسوية - تمت تسوية التناقضات بين صياغة الرموز والتحقق من الرموز من خلال الاستعانة بمسؤول ترميز أو محكم خارجي أعلى رتبة.

- بعد ترميز الإجابات المتبقية واجتياز إجراءات مراقبة الجودة، أُضيفت الإجابات إلى الملف الرئيسي للمساعدة في عملية الترميز الآلي عن طريق إنشاء مطابقات محتملة إضافية للإجابات المكتوبة المستقبلية.

خاتمة

تطوير أنظمة ترميز فعالة للإجابات المكتوبة من المهام الأكثر إرهاقًا وصعوبة في التعداد السكاني. لكن لحسن الحظ، يمكن تقليل عبء أعمال الترميز وتخفيض ميزانيتها من على كاهل مكتب الإحصاء الوطني (NSO) من خلال التخطيط السليم والاستخدام الفعال للتكنولوجيا. إن نظام الترميز الآلي المصمم بشكل صحيح، والذي يتضمن بعض مكونات الترميز البشرية، من شأنه أن يؤدي إلى جعل العملية تتم آليًا إلى أقصى حد ممكن من دون المساس بالجودة.

المراجع

Amón, I., Moreno, F., and Echeverri, J., "Algoritmo fonético para detección de cadenas de texto duplicadas en el idioma español," *Revista Ingenierías, Universidad de Medellín*, 11(20), 127–138, 2012.

Bethlehem, J., "Applied Survey Methods: A Statistical Perspective," Wiley series in survey methodology, Hoboken, N.J., Wiley, 2009.

Campanelli, P., Thomson, K., Moon, N., and Staples, T., "The Quality of Occupational Coding in the United Kingdom," In Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D., eds., *Survey measurement and process quality,* pp. 437–453, New York, Wiley, 1997.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R., Survey methodology 2nd ed., Wiley series in survey methodology, Hoboken, N.J., Wiley, 2009.

Lyberg, L., and Kasprzyk, D., "Some aspects of post-survey processing," In Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. eds., *Survey measurement and process quality,* pp. 353–370), New York, Wiley, 1997.

Rea, L., and Parker, R., "Designing and Conducting Survey Research: A Comprehensive Guide," fourth ed., San Francisco, CA, Jossey-Bass, a Wiley brand, 2014.

Russell, U.S. Patent No. 1261167, 1918.

Saris, W., and Gallhofer, I., "Design, Evaluation, and Analysis of Questionnaires for Survey Research," second ed., Wiley series in survey methodology, Hoboken, New Jersey, Wiley, 2014.

U.S. Census Bureau, American Community Survey, Design and Methodology, Washington, D.C., U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2009.

Wolf, C., Joye, D., Smith, T., and Fu, Y., eds., "The Sage Handbook of Survey Methodology," Los Angeles, Sage Publications, 2016.

يتم نشر سلسلة مواضيع مختارة في التعدادات السكانية الدولية (STIC) بواسطة برامج دولية داخل شعبة السكان التابعة لمكتب الإحصاء الأمريكي. ترعى الوكالة الأمريكية للتنمية الدولية إنتاج سلسلة مواضيع مختارة في التعدادات السكانية الدولية (STIC)، فضلاً عن الدعم المزدوج الموجه إلى المنظمات الإحصائية التي تُثري خبرات المؤلفين. كما يتعاون صندوق الأمم المتحدة للسكان، فيما يتعلق بالمحتوى والنشر، بما يضمن وصول سلسلة مواضيع مختارة في التعدادات السكانية الدولية (STIC) إلى جمهور أوسع نطاقًا.