مـركـز الإحـصـاء
STATISTICS CENTRE

عملية التشغيل والتحليل للبيانات الإحصائية وفقاً للنموذج المعياري لإدارة العمليات الإحصائية

أدلة المنهجيات الجودة – رقم الدليل (23)

قائمة المحتويات

1. المقدّمة..... 3
2. النموذج المعياري لإدارة العمليات الإحصائية.....3
2.1. مرحلة عملية التشغيل.....5
2.2. مرحلة التحليل.....10
3. خصوصيّات كلّ نوع من أنواع المسح 12
3.1. المسوح الأسرية.....12
3.2. مسوح المنشآت الاقتصادية.....13
4. المرجع.....16

1. المقدمة

تساهم المبادئ التوجيهية الخاصة بمعالجة وتدقيق بيانات المسوح ومراقبة جودتها، في تعزيز الجوانب المنهجية لتطوير وإنتاج مؤشرات جديدة، وذلك من خلال اتباع أفضل المعايير والممارسات الدولية.
بالنسبة لجميع العمليات الإحصائية، سواء بالنسبة للإحصاءات القائمة على المسح أو الإحصاءات المعتمدة على السجلات الإدارية ، يمكن استخدام النموذج المعياري لإدارة العمليات الإحصائية (GSBPM) لوصف المراحل المختلفة والمراحل الفرعية للعمل الإحصائي
تقدم هذه الوثيقة النموذج المعياري لإدارة العمليات الإحصائية (GSBPM) وصف للعمليات الاحصائية مع التركيز على مراحلها المتعلقة بأنشطة معالجة البيانات وتحليلها، يتم تقديمها في القسم الثاني. يركز القسم الثالث بشكل أكثر تحديدًا على نوعين من المسوح، المسوح الخاصة بالأسر والمسوح المتعلقة بإحصاءات الأعمال. لاحظ أن جميع العمليات الفرعية بما في ذلك القسم الخامس والسادس، تعتمد على وثيقة التصميم. والذي بدوره، يعتمد على احتياجات المستخدم.
كما تم ذكره سابقا، تركز هذا الدليل على توجيه القارئ خلال المراحل المختلفة للمعالجة والتحليل، مع أمثلة من الإحصائيات القائمة على المسح. من المهم، أن نتذكر أن كلاً من الإحصائيات القائمة على المسح والإحصاءات القائمة على الإدارة تتبع بشكل أساسي نفس العملية وفقًا لـ GSBPM. بعض الاختلافات الواضحة هي، على سبيل المثال أننا لا نحتاج إلى آلية سحب عينات أو استمارة للبيانات الإدارية.
في هذه الوثيقة يتم فقط التركيز على عمليتين من ثماني عمليات، وبالكاد نذكر المرحلة الأولية من عملية الإنتاج الإحصائي والتي تهدف دائمًا إلى التأكد من فهم احتياجات المستخدم وما سيتم استخدام النتيجة الإحصائية النهائية من أجله. فمن المستحيل إنتاج معلومات إحصائية عالية الجودة من دون هذه المعلومات.

2. النموذج المعياري لإدارة العمليات الإحصائية

يُعتبر النموذج المعياري لإدارة العمليات الإحصائية (GSBPM)[1] المُعتمد من قبل المراكز الإحصائية للدول الأكثر تقدّمًا في مجال الإحصاء (الجدول 1)، مفيداً لوصف جميع مراحل إنتاج الإحصاءات ونشرها. ويهدف هذا النموذج إلى توجيه عملية تخطيط المسوح وغيرها من العمليات الإحصائية، من خلال المراجعة المنهجية لجميع العمليات وإدارة التدفق بدءًا من المراحل الأولى للإعداد للنشر والتوثيق والأرشفة.

العمليات الاحصائية هي مجموعة من الأنشطة والمهام ذات الصلة والمنظمة لتحويل البيانات المدخلة إلى معلومات إحصائية. في سياق GSBPM، يقوم مركز الإحصاء – أبوظبي بعمليات الأعمال الإحصائية لإنشاء إحصاءات رسمية لتلبية احتياجات المستخدمين. قد يكون ناتج العملية مجموعة مختلطة من المنتجات مادية أو رقمية التي تقدم الإحصاءات والبيانات الوصفية بطرق مختلفة، مثل المنتجات الإحصائية والمنشورات والخرائط والخدمات الإلكترونية، من بين غيره.

تستخرج هذه الوثيقة الأقسام المرتبطة بالعمليات الإحصائية من مراحل المعالج والتحليل.

[1] يستند هذا القسم إلى شعبة الإحصاءات التابعة للجنة الأمم المتحدة الاقتصادية لأوروبا
(https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1) والتي تم توفيرها لاستخدامها من قبل جميع المراكز الإحصائية. ونظرا لأن النموذج المعياري لإدارة العمليات الإحصائية (GSBPM) يعد معيارًا إحصائيًا، فقد تم الحفاظ على التعريفات المرتبطة بالمراحل الرئيسية والمراحل الفرعية على النحو الذي ترد فيه في المستند الأصلي. وقد تمّ ترخيص المستند الأصلي المعني بالنموذج المعياري لإدارة العمليات الإحصائية (بموجب الترخيص الدولي Creative Commons Attribution 4.0) وهو منسوب إلى لجنة الأمم المتحدة الاقتصادية لأوروبا نيابة عن المجتمع الإحصائي الدولي.

الجدول 1: النموذج المعياري لإدارة العمليات الإحصائية

عمليات الشمول							
تحديد الاحتياجات	**التصميم**	**البناء**	**الجمع**	**عملية التشغيل**	**التحليل**	**النشر**	**التقييم**
1.1 تحديد الاحتياجات	2.1 تصميم المخرجات	3.1 إعادة استخدام أو بناء أدوات الجمع	4.1 إنشاء الإطار واختيار العينة	5.1 دمج البيانات	6.1 إعداد مسودة المخرجات	7.1 تحديث أنظمة المخرجات	8.1 جمع مدخلات التقييم
1.2 التشاور وتأكيد الاحتياجات	2.2 تصميم الأوصاف المتغير	إعادة استخدام أو بناء مكونات المعالجة والتحليل	4.2 الإعداد لعملية الجمع	5.2 التصنيف والترميز	6.2 التحقق من صحة المخرجات	7.2 إنتاج مخرجات النشر	8.2 إجراء التقييم
1.3 تحديد أهداف المخرجات	2.3 تصميم عملية الجمع	إعادة استخدام أو بناء مكونات النشر	4.3 إجراء عملية الجمع	5.3 المراجعة والتحقق من صحة البيانات	6.3 تفسير وشرح المخرجات	7.3 إدارة إصدار منتجات النشر	8.3 الاتفاق على خطة العمل
1.4 تحديد المفاهيم	2.4 تصميم الإطار والعينة	3.4 تكوين مهام سير العمل	4.4 الانتهاء من عملية الجمع	5.4 التحرير والتعامل مع القيم المفقودة	6.4 تطبيق مراقبة الإفصاح عن البيانات	7.4 الترويج لمنتجات النشر	
1.5 التحقق من توفر البيانات	2.5 تصميم العملية التشغيلية والتحليل	3.5 اختبار أنظمة الإنتاج		5.5 اشتقاق متغيرات ووحدات جديدة	6.5 الانتهاء من ملفات البيانات	7.5 إدارة دعم المستخدم	
1.6 إعداد وتقديم حالة العمل	2.6 تصميم أنظمة الإنتاج وسير العمل	3.6 اختبار إجراءات الإنتاج الإحصائية		5.6 حساب الأوزان			
		3.7 استكمال نظام الإنتاج		5.7 حساب المجاميع			
				5.8 الانتهاء من ملفات البيانات			

بحسب النموذج تشير المراحل الملونة بالازرق الى (المستوى الأول) والعمليات الفرعية الملونة بالوردي (المستوى الثاني). لاحظ أن مرحلتي "عملية التشغيل" و "التحليل" يمكن أن تكون متكررة ومتوازية. يمكن أن يكشف التحليل عن فهم أوسع للبيانات، مما قد يوضح بوجود حاجة إلى معالجات إضافية.

2.1 مرحلة عملية التشغيل

الجدول 2: النموذج المعياري لإدارة العمليات الإحصائية "مرحلة عملية التشغيل"

عملية التشغيل					
5.3 المراجعة والتحقق من صحة البيانات	5.4 التحرير والتعامل مع القيم المفقودة	5.5 اشتقاق متغيرات ووحدات جديدة	5.6 حساب الأوزان	5.7 حساب المجاميع	5.8 الانتهاء من ملفات البيانات

تصف مرحلة "عملية التشغيل" كيفية معالجة البيانات المدخلة وإعدادها للتحليل، وهي تتكون من عمليات فرعية تدمج البيانات المدخلة وتصنّفها وتفحصها وتحولها بهدف التمكن من تحليلها ونشرها على شكل مخرجات إحصائية. وفيما يتعلّق بالمخرجات الإحصائية المنتجة بشكل منتظم، تكرر هذه المرحلة في كل مرة.

يمكن أيضًا أن تبدأ الأنشطة ضمن مرحلتي " عملية التشغيل" و "التحليل" قبل اكتمال مرحلة "الجمع". مما يتيح ذلك من تجميع النتائج المؤقتة، حيث يكون عامل التوقيت مهمًا للمستخدمين، ويزيد من الوقت المتاح للتحليل.

تُقسم مرحلة "المعالجة" إلى ثماني عمليات فرعية (الجدول 2)، قد تكون متتالية من اليمين إلى اليسار، أو حتى متوازية، كما يمكن أن تكون مكرّرة. وتاليا لمحة عن العمليات الثمانية الفرعية المرتبطة بأنشطة المعالجة والمطابقة وحساب أوزان بيانات المسح:

المرحلة 5.1 دمج البيانات

في هذه العملية الفرعية يتم دمج البيانات من مصدر واحد أو أكثر، وهي المرحلة التي تُجمع فيها نتائج العمليات الفرعية في مرحلة "الجمع". ويمكن أن تكون البيانات المدخلة من مصادر خارجية أو داخلية مختلفة، وبواسطة أدوات جمع مختلفة، بما في ذلك البيانات المستخرجة من السجلات الإدارية ومصادر البيانات غير الإحصائية الأخرى. كما تشمل العملية الفرعية مواءمة المخرجات، أو إنشاء مخرجات جديدة، تتوافق مع مصادر البيانات المختلفة، بحيث تكون النتيجة عبارة عن مجموعة مترابطة من البيانات، وتشمل عملية دمج البيانات ما يلي:

- دمج البيانات التي تمّ الحصول عليها من مصادر متعددة كجزء لإعداد الإحصاءات المتكاملة مثل الحسابات القومية.
- دمج البيانات الجغرافية المكانية والبيانات الإحصائية أو غيرها من البيانات غير الإحصائية.
- تجميع البيانات بهدف زيادة العدد الفعال للمشاهدات لرصد بعض الظواهر.
- القيام بإجراءات مطابقة أو ربط البيانات السجلية، بهدف ربط البيانات الجزئية أو البيانات الكلية من مختلف المصادر.
- دمج البيانات ومن ثمّ التخفيض أو الاستبدال.

- تحديد الأولويات في الحالات التي يحتوي فيها مصدران أو أكثر على بيانات لنفس المتغيّر مع احتمال اختلاف القيم.

لدى مركز الإحصاء – أبوظبي (SCAD) نموذج تدفق البيانات الذي يغطي العملية الفرعية المتعلقة بتكامل البيانات، انظر الى الدليل رقم (19) للمرجعية. لاحظ أن كل من بيانات المسوح والبيانات الإدارية تتم معالجتها بطريقة مماثلة.

المرحلة 5.2 التصنيف والترميز

تقوم هذه العملية الفرعية بتصنيف وترميز البيانات. فعلى سبيل المثال؛ قد تحدّد إجراءات الترميز التلقائية (أو المكتبية) رموزًا عددية للإجابات النصية، وفقًا لتصنيف إحصائي محدد مسبقا، وذلك بهدف تسهيل عملية جمع البيانات ومعالجتها.

وتحتوي بعض الأسئلة في استمارات المسوح على فئات إجابة مرمّزة، فيما يتمّ ترميز بعضها الآخر بعد الجمع من خلال عملية آلية (والتي قد تطبّق تقنيات التعلم الآلي، مثل ترميز مبيعات التجزئة) أو عملية يدوية تفاعلية.

الترميز؛ تشير عملية الترميز إلى وضع قيمة عددية للاستجابات، بهدف تسهيل عملية جمع البيانات. فعندما يتم تحديد نظام الترميز، يجب أن يكون الهدف هو تصنيف الإجابات ضمن مجموعة ذات معنى من الفئات الشاملة والحصرية المتبادلة التي تبرز النمط الأساسي للإجابات، وبالنسبة لبعض الأسئلة، قد يكون الترميز مباشرًا (كالحالة الاجتماعية)، وأمّا في الحالات الأخرى مثل؛ الموقع الجغرافي والنشاط الصناعي والمهنة، فقد يكون هنالك دليل ترميز معياري، ومع ذلك، تبقى بعض الأسئلة من دون نظام ترميز معياري، مما يجعل وضع نظام ترميز جيد ذو مهمّة لا يستهان بها، ويجب أن يكون دليل الترميز متسقًا ومنطقيًّا، مع الأخذ في الاعتبار مدى تفصيل الترميز المطلوب، وذلك على ضوء الغرض من المسح وجداول المخرجات أو تحليل البيانات الواجب القيام به.

من الأفضل البدء بقائمة موسعة، نظرًا لأن عددًا قليلاً جدًا من الفئات يمكن أن يكون مضللًا، ويمكن أن تكون فئة أخرى كبيرة غير مفيدة. يمكن دائمًا تصغير الفئات، ولكن قد يكون من الصعب تقسيمها إلى الفئات الأصلية بعد ذلك. رميز البيانات هو تحويل الاجابات إلى تنسيق يمكن قراءته آليًا. باستخدام طرق التجميع المعتمدة على الكمبيوتر، يتم الترميز في وقت جمع البيانات.

التدقيق هو عملية فحص البيانات، من أجل تحديد البيانات المفقودة أو غير الصحيحة أو غير المتسقة، والتي تشير إلى سجلات البيانات التي من المحتمل أن تكون خاطئة، وعادةً ما تحدد عملية تدقيق الأخطاء غير المتعلقة بالمعاينة، والناتجة عن أخطاء القياس (الاستجابة) أو عدم الاستجابة أو المعالجة، ويمكن أن تظهر عملية التدقيق في عدد من المراحل خلال تنفيذ المسح، من خلال الفحوصات الأولية البسيطة، التي يقوم بها الباحثون في الميدان، إلى عمليات التحقق الآلية الأكثر تعقيدًا، والتي تتم من خلال برامج التدقيق الالي المعدة مسبقا للتحقق البيانات.

تألّف عملية التدقيق من ثلاث فئات رئيسية هي الصلاحية والاتساق وتدقيق التوزيع، مع الإشارة إلى أنّه يتمّ تطبيق كلّ من التحقق والاتساق لكلّ استمارة على حدة.

- يتم التحقق من الصلاحية من خلال تدقيق صياغة الأجوبة، كالتحقق من الأحرف غير الرقمية التي تم الإبلاغ عنها في الحقول الرقمية، كما يتم التحقق من القيم المفقودة، أو أن البيانات المرمزة تقع ضمن نطاق القيم المسموح بها .

- لتحقق من الاتساق ومراعات العلاقات بين الأسئلة، من خلال الاستناد إلى العلاقات المنطقية، أو القانونية، أو المحاسبية، أو الهيكلية بين الأسئلة أو أجزائها.
- يتمّ إجراء التحقق من توزيع البيانات، بالنظر للبيانات خلال الاستمارات، بهدف محاولة تحديد القيم المتطرفة فيما يتعلق بتوزيع البيانات.

التدقيق خلال مرحلة جمع البيانات غالبًا ما يُشار إليها على أنّها عمليات تدقيق ميدانية، وغالبًا ما تتألّف من عمليات تدقيق الصلاحية، بالإضافة إلى عمليات بسيطة لتدقيق الاتساق.

تتمّ عمليات التدقيق الأكثر شمولًا وتعقيدًا كمرحلة تدقيق وتعويض منفصلة بعد مرحلة جمع البيانات.

فأثناء عملية جمع البيانات، يمكن إجراء عمليات التدقيق من قبل المدخلين، أو بشكل تلقائي عن طريق برامج الحاسوب، أو من خلال تطبيق الحاسوب في حالة الاستعانة بطرق الجمع من خلالها، وبشكل عام، يتمّ الحدّ من عملية التدقيق أثناء جمع البيانات، نظرًا لأنّ عملية الاستجابة للتدقيق تبطئ سرعة عملية جمع البيانات، وعادةً ما تكون عمليات التدقيق في هذه المرحلة من معالجة البيانات، محصورة بتدقيق الصلاحية وعمليات تدقيق بسيطة للاتساق

في هذا الدليل، لن نتعرض الى التدقيق المرتبط بالمستوى الكلي (مجاميع البيانات أو الإحصائيات).

المرحلة 5.3 المراجعة والتحقق من صحة البيانات

تجرى هذه العملية الفرعية على البيانات لتحديد المشاكل والأخطاء والتناقضات المحتملة، مثل القيم المتطرفة وعدم الاستجابة والخطأ في الترميز، كما يمكن الإشارة إليها على أنّها عملية تحقق من البيانات المدخلة، ويمكن القيام بهذه العملية بشكل متكرّر والتحقق من البيانات، بالمقارنة مع قواعد التدقيق المحددة مسبقًا ضمن ترتيب محدّد في أكثر الأوقات، كما يمكنها أن تضع علامة على البيانات التي تتطلّب التحقق منها أو تدقيقها، سواء بشكل تلقائي أو يدوي، ويمكن تطبيق عملية المراجعة والمطابقة على البيانات المستخرجة من أي نوع من أنواع المصادر قبل إدماجها وبعده، بالإضافة إلى البيانات المنسوبة إلى العملية الفرعية 5.4 (التدقيق والتعويض)، في حين يُعتبر التحقق جزءًا من مرحلة "المعالجة"، ولكن من الناحية العملية قد تتم بعض عناصر التدقيق جنباً إلى جنب مع أنشطة الجمع، لا سيما في بعض الأنماط، مثل الجمع بمساعدة برامج الحاسوب، وبينما تهتمّ هذه العملية الفرعية بالكشف عن الأخطاء الفعلية أو المحتملة وتحديد موقعها، نجد أن أنشطة التصحيح التي تعدّل البيانات فعليًّا تتم في العملية الفرعية 5.4 (التدقيق والتعويض).

يعد التحقق من صلاحية البيانات نشاطًا للتحقق ممّا إذا كانت تركيبة معيّنة من القيم جزءا من مجموعة التركيبات المقبولة أم لا، إذ يقيّم هذا الإجراء منطقية البيانات، لا تضمن النتائج الإيجابية صحة البيانات، إنّما تضمن النتائج السلبية عدم صحتها.

المرحلة 5.4 التحرير والتعامل مع القيم المفقودة

في الحالات التي تُعتبر فيها البيانات غير صحيحة، أو ناقصة أو غير موثوقة أو غير محدّثة، يجب إدخال قيم جديدة، أو حذف البيانات القديمة ضمن هذه العملية الفرعية. يغطي المصطلحان التحرير والتعويض مجموعة متنوعة من الأساليب للقيام بذلك، غالبًا تكون باستخدام نهج قائم على القواعد. وتتضمن الخطوات المحددة عادةً ما يلي:

- تحديد ما إذا كان سيتم إضافة البيانات أو تغييرها؛
- اختيار الطريقة الواجب اتباعها؛

- إضافة/تغيير قيم البيانات؛
- إعادة تدوين قيم البيانات الجديدة في مجموعة البيانات والإشارة إلى ان البيانات تمّ تغييرها؛
- إنتاج البيانات الوصفية المتعلقة بعمليتي التدقيق والتعويض.

وضع مركز الإحصاء – أبو ظبي تدابير عديدة لضمان بيانات ذات دقة ومعاير جودة عالية، ويمكن تلخيص مراحل التدقيق على الشكل التالي:

- التدقيق الميداني: تشكّل هذه العملية أولى مراحل التدقيق، حيث يقوم الباحثون والمراقبون بمراجعة الاستمارات المكتملة قبل تسليمها إلى المكتب.
- التدقيق المكتبي: يراجع فريق التدقيق الاستمارات المكتملة، ويقوم بإبلاغ الفريق الميداني بشأن الأخطاء التي تتطلّب تصحيحاً في الميدان، بالإضافة إلى الأخطاء المتكررة من قبل الباحثين. بعد ذلك، يتمّ ترميز الاستمارات بالاستناد إلى التصنيفات المعتمدة، وتحميلها في قاعدة البيانات من خلال نظام إدخال البيانات.
- التدقيق الإلكتروني: بمجرّد إدخال البيانات إلى النظام، يتمّ التحقق منها تلقائيًّا للإبلاغ عن أخطاء البيانات الميدانية المتسلسلة، وللتحقق من القيم الشاذة، يتم تطبيق القواعد التالية؛ التدقيق الهيكلي، والاتساق وصحة المتغيرات، كما يتمّ استخدام عمليات التدقيق الآلي للتحقق من مجاميع القيم، وصحة المتغيرات ذات الصلة، بالإضافة إلى عمليات التحقق المنطقية.
- عمليات التحقق الأخرى: تتألّف هذه العمليات بشكل أساسي من تحليل تناسق السلاسل الزمنية للبيانات الجزئية ونتائج المسح.

طرق التعويض

التعويض هو عملية استبدال البيانات المفقودة والقيم غير الصحيحة والقيم المتطرفة بقيم تقديرية، باستخدام طرق التعويض المختلفة، ومن ثم تحليل البيانات الكاملة باعتبار القيمة المحتسبة على أنّها قيمة فعلية.

تشكل البيانات المفقودة وأخطاء البيانات والقيم المتطرفة تحديات شائعة في مجال التحليل الإحصائي، والتي يمكن مصادفتها خلال عملية التأكد من صحة البيانات وتحليلها. ملاحظة، الخبير في الموضوع هو الأنسب غالبًا لتقييم القيم المتطرفة. عندما يتعلق الأمر بتحديد آلية التعامل مع القيم المتطرفة، يجب التعاون بينه وبين فريق المنهجية.

فالبيانات المفقودة: هي الحالة التي تكون فيها قيمة واحدة او أكثر من المشاهدات في متغير او عدد من المتغيرات ضمن قاعدة البيانات غير متوفرة. بينما البيانات غير الصحيحة هي الحالة التي يكون فيها حدوث قيمة غير صحيحة في واحدة أو أكثر من المشاهدات في متغير أو مجموعة من المتغيرات ضمن قاعدة البيانات أمّا القيم المتطرفة؛ فهي الحالة التي تكون فيها قيمة المشاهدة تختلف معنويا عن باقي المشاهدات المسجلة ضمن قاعدة البيانات.

يعتمد اختيار طريقة التعويض المناسبة على جوانب متعددة، مثل؛ حجم قاعدة البيانات، ونوع المتغير ومعدل البيانات المفقودة، والبيانات غير الصحيحة والقيم المتطرفة، وتوزيع المتغير والسلاسل الزمنية، والاتساق ومعدل النمو وتوافر البيانات السابقة، وضوضاء البيانات وتصنيفها.

في معظم المسوح، يتعيّن على الفرد مواجهة مشكلة البيانات المفقودة، وعادة ما يتمّ التمييز بين نوعين من البيانات المفقودة وهي؛ عدم استجابة الوحدة (على سبيل المثال منشاة/ أسرة)، وعدم استجابة العنصر. يتمّ تطبيق التعويض في الحالة الأخيرة، ويتعيّن إجراء إعادة الترجيح لعدم استجابة الوحدة، كما يشير عدم استجابة

الوحدة إلى التعذر عن جمع أي من المعلومات عن بعض وحدات المسح، فيما يشير عدم استجابة العنصر إلى البيانات المفقودة في الاستمارات التي يتمّ إرجاعها.

عدم استجابة الوحدة: تكمن الممارسة الإحصائية الشائعة لتصحيح عدم استجابة الوحدة في تغيير الأوزان بهدف تعويض عدم استجابة الوحدة.

عدم استجابة العنصر: في حالة عدم الاستجابة الفنية أو الحساسة، قد يتم الحدّ منها إذا تم تحديد الشخص الذي لديه معرفة بالموضوع أو مسؤولية أعلى، يجب أن يستجيب بالنيابة عن المنشأة.

التعويض بناء على الخبرات السابقة

يتمّ تعويض القيمة المفقودة (والقيم غير الصحيحة) مع القيمة المتوقعة أو المقدرة المقدمة من قبل الخبراء في الموضوع. تتألّف هذه الطريقة من استبدال القيم المفقودة بالقيم التي يحددها الخبراء في الموضوع باستخدام إجراء يدوي أو بعض القواعد العامة. يوصى بهذه الطريقة فقط في حالة العدد القليل جدًا من القيم المفقودة، ويكون للخبراء في الموضوع توقعات أو تقديرات لكل من القيم المفقودة التي سيتم استخدامها للتعويض

فعلى سبيل المثال، القيم المفقودة في أجور العمل أو واجبات مهام محددة. يمكن للخبراء في الموضوع الحصول على المعلومات، من خلال تجربتهم في السوق، واقتراح البيانات التي يجب تعويضها للمتغيرات المفقودة.

المرحلة 5.5 اشتقاق متغيّرات ووحدات جديدة

تشتق هذه العملية الفرعية المتغيرات والوحدات التي لم يتمّ طلبها بشكل صريح عند جمع البيانات، ولكنها ضرورية لتوفير المخرجات المطلوبة، حيث تشتق هذه العملية متغيرات جديدة، من خلال تطبيق صيغ حسابية على متغيّر واحد أو أكثر من المتغيرات الموجودة في مجموعة البيانات، أو تطبيق نماذج افتراضية مختلفة، وقد يتطلب الأمر تكرار هذا النشاط نظرًا لأن بعض المتغيرات المشتقة بحدّ ذاتها تعتمد على متغيرات مشتقة أخرى. وبالتالي فانه من الضروري ضمان اشتقاق المتغيرات بالترتيب الصحيح، حيث يمكن اشتقاق الوحدات الجديدة من خلال تجميع أو تقسيم بيانات عدد من الوحدات، أو من خلال طرق تقدير أخرى مختلفة. من الأمثلة على ذلك، اشتقاق بيانات الأسر، حين تكون وحدات الجمع هي الأفراد، أو المؤسسات حين تكون وحدات الجمع هي الوحدة القانونية (المؤسسة).

المرحلة 5.6 حساب الأوزان

تنشئ هذه العملية الفرعية أوزانًا لوحدات بيانات سجلات وفقًا لمنهجية تصميم العينة لكلّ مسح، فعلى سبيل المثال؛ يمكن استخدام الأوزان في تجميع البيانات لتمثّل المجتمع المستهدف (بالاعتماد على بيانات لعينة مسح احصائي)، أو لتعديل حالات عدم الاستجابة في عمليات التعداد الكلي، وفي الحالات الأخرى، قد يكون من الضروري استخدام الأوزان لأغراض معايرة البيانات (normarization purpose)، كما قد يشمل تصحيح الأوزان لمؤشرات المقارنة (مثل إجمالي عدد السكان المعروف).

المرحلة 5.7 حساب المجاميع

تنشئ هذه العملية الفرعية بيانات مجمعة وإجماليات المجتمع من البيانات الجزئية، أو التجميعات ذات المستوى الأدنى، وهي تشمل جمع البيانات للسجلات التي تشترك في خصائص معينة (مثل تجميع البيانات حسب التصنيفات الديموغرافية أو الجغرافية)، وتحديد مقاييس المتوسط والتشتت، وتطبيق الأوزان المُشار إليها في

العملية الفرعية 5.6 احتساب الأوزان) لاشتقاق المجاميع المناسبة، وفي حالة المخرجات الإحصائية التي تستخدم مسوح العينة، يمكن احتساب أخطاء المعاينة المقابلة للمؤشرات المعنية ضمن إطار هذه العملية الفرعية.

المرحلة 5.8 الانتهاء من ملفات البيانات

في هذه العملية الفرعية؛ يتم تجميع نتائج العمليات الفرعية الأخرى ضمن ملفّ بيانات (غالبًا ما تكون بيانات كلية) والذي يستخدم كمدخل لمرحلة "التحليل". وفي بعض الأوقات، قد يكون هذا ملفًا اوليا وليس نهائيًا ، حيث تكون هناك ضغوط ومتطلبات لإنتاج كل من التقديرات الأولية والنهائية في وقت واحد.

2.2 مرحلة التحليل

الجدول 3: مرحلة التحليل من مراحل النموذج المعياري لإدارة العمليات الإحصائية

التحليل				
6.1	6.2	6.3	6.4	6.5
إعداد مسودة المخرجات	التحقق من صحة المخرجات	تفسير وشرح المخرجات	تطبيق مراقبة الإفصاح عن البيانات	الانتهاء من ملفات البيانات

في هذه المرحلة، يتمّ إنتاج المخرجات الإحصائية وفحصها بالتفصيل، وتشمل إعداد المحتوى الإحصائي (بما في ذلك التعليقات والملاحظات الفنية وغيرها)، وضمان ملاءمة المخرجات لتحقيق الأهداف قبل نشرها للمستخدمين. بالإضافة إلى ذلك، تشمل هذه المرحلة العمليات والأنشطة الفرعية التي تمكّن المحللين الإحصائيين من فهم البيانات والإحصاءات المنتجة، كما يمكن استخدام مخرجات هذه المرحلة كمدخلات للعمليات الفرعية الأخرى (على سبيل المثال؛ تحليل المصادر الجديدة كمدخلات لمرحلة "التصميم"). وبالنسبة للمخرجات الإحصائية التي يتم إنتاجها بانتظام، تطبق هذه المرحلة في كل تكرار، كما تُعتبر مرحلة "التحليل" والعمليات الفرعية لها عامّة لجميع المخرجات الإحصائية، بغض النظر عن مصدر البيانات وكيفية الحصول عليها.

تنقسم مرحلة "التحليل" إلى خمس عمليات فرعية (الجدول 3)، والتي عادة ما تكون متسلسلة من اليمين إلى اليسار، كما يمكن أن تحدث بالتوازي وبوتيرة متكررة، والعملية الفرعية الوحيدة التي تشير إلى أنشطة التحقق من صلاحية بيانات المسح هي العملية الفرعية 6.2.

المرحلة 6.1 إعداد مسودة المخرجات

هذه العملية الفرعية يتم من خلالها تحويل البيانات من العمليات الفرعية 5.7 (حساب المجاميع) و 5.8 (الانتهاء من ملفات البيانات) إلى مخرجات إحصائية مثل القيم الإحصائية (التقديرات) ، والمؤشرات ، والإحصاءات المعدلة موسمياً ، على سبيل المثال ، الاتجاه ، الدورة ، الموسمية والمكونات غير المنتظمة ، ومقاييس الوصول، وما إلى ذلك ، فضلاً عن تسجيل خصائص الجودة (على سبيل المثال ، معاملات التغير). يمكن تضمين إعداد الخرائط ومخرجات نظم المعلومات الجغرافية والخدمات الجغرافية الإحصائية لتعظيم القيمة والقدرة على تحليل المعلومات الإحصائية. هذه هي أيضًا المرحلة التي يتم فيها اختبار البيانات والتخطيط لتقييم الجودة النهائية الذي يجب إجراؤه على مستوى إجمالي ، على سبيل المثال ، التدقيق الكلي والافصاح عن البيانات. في هذه المرحلة، يجب كتابة مسودة تقرير الجودة النهائي (ليس لاحقًا).

المرحلة 6.2 التحقق من صحة المخرجات

تتألّف العملية الفرعية 6.2 من مطابقة جودة المخرجات المنتجة وفقًا لإطار الجودة العام والتوقعات، وقد تشمل أنشطة المطابقة ما يلي:

- التحقق من أنّ معدلات التغطية ونسب الاستجابة مستوفية للمتطلبات.
- مقارنة الإحصاءات بالدورات السابقة (إذا أمكن).
- التحقق من وجود البيانات الوصفية والبيانات الوصفية لعملية جمع البيانات ومؤشرات الجودة وتماشيها مع التوقعات.
- التحقق من الاتساق الجغرافي المكاني للبيانات.
- مقابلة الإحصاءات مع البيانات الأخرى ذات الصلة (على الصعيدين الداخلي والخارجي).
- تقصي وجود عدم اتساق في الإحصاءات.
- إجراء عملية التدقيق الكلي.
- مطابقة الإحصاءات مع التوقعات.

المرحلة 6.3 تفسير وشرح المخرجات

يتم خلال هذه لعملية الفهم المتعمق للمخرجات. يستخدمون هذا الفهم لتفسير الإحصائيات وشرحها من خلال تقييم مدى جودة الإحصاءات التي تعكس توقعاتهم الأولية، وعرض الإحصاءات من جميع وجهات النظر باستخدام أدوات ووسائل مختلفة ، وإجراء تحليلات إحصائية متعمقة مثل تحليل السلاسل الزمنية والاتساق و تحليل المقارنة ، تحليل المراجعة (تحليل الفروق بين التقديرات الأولية والمنقحة) ، تحليل التباينات (مثال التناقضات في احصاءات المرأة) ، إلخ.

المرحلة 6.4 تطبيق مراقبة الإفصاح عن البيانات

تضمن هذه العملية الفرعية أن البيانات (والبيانات الوصفية) التي سيتم نشرها لا تنتهك القواعد المناسبة بشأن السرية وفقًا لسياسات وقواعد المؤسسة، أو للمنهجية الخاصة بالعملية التي تم إنشاؤها في العملية الفرعية 2.5 (تصميم العملية التشغيلية والتحليل). قد يشمل ذلك عمليات التحقق من الإفصاح الأولي والثانوي، فضلاً عن تطبيق تقنيات التحفظ على البيانات او اطلاقها. قد تختلف درجة وطريقة التحكم في الإفصاح الإحصائي باختلاف أنواع المخرجات. على سبيل المثال، ستكون المنهجية المستخدمة لمجموعات البيانات التفصيلية للأغراض البحثية مختلفًا عن ذلك الخاص بالجداول المنشورة والمخرجات النهائية للإحصاءات الجغرافية المكانية أو التصورات على الخرائط.

المرحلة 6.5 الانتهاء من ملفات البيانات

ضمن هذه العملية الفرعية أن الإحصائيات والمعلومات المرتبطة بها مناسبة للغرض وتصل إلى مستوى الجودة المطلوب المحدد في وثيقة التصميم وبالتالي فهي جاهزة للاستخدام. وتشمل:

- استكمال فحوصات التناسق.
- تحديد مستوى الإطلاق وتطبيق المحاذير.
- تجميع المعلومات الداعمة، بما في ذلك التفسير والتعليق والملاحظات الفنية والإنجازات وتدابير عدم الدقة وأي بيانات وصفية ضرورية أخرى.

- إنتاج المستندات الداخلية الداعمة.
- إجراء مناقشة ما قبل النشر مع خبراء الموضوع الداخليين المناسبين.
- ترجمة المخرجات الإحصائية في الدول ذات النشر المتعدد اللغات.
- اعتماد المحتوى الإحصائي للنشر.
- إنتاج وإنهاء تقرير الجودة. يتم ذلك من خلال المنهجية والموضوع بالتعاون (الموضوع مسؤول).

3. خصوصيّات كلّ نوع من أنواع المسح

يصف هذا القسم الجوانب المنهجية لمعالجة بيانات المسح، والتحقق من صحتها ومراقبة جودتها، فيما يتعلّق بثلاثة أنواع من المسوح وهي الأسرية والاقتصادية والحكومية.

3.1 المسوح الأسرية

فيما يلي وصف طرق المعالجة والتحقق من الصحة ومراقبة الجودة المذكورة أدناه المطبقة على المسوح الأسرية.

بالنسبة لمسوح الأسر المعيشية، عادة ما تحتوي استمارة المسح على قسم للبيانات على مستوى الأسرة ، وقسم آخر لكل فرد في الأسرة (يتم تكرارها عدة مرات مثل عدد الأفراد في الأسرة). هذا يعني أن معالجة البيانات تتم باستخدام تجميع نوعين من الوحدات الإحصائية: الأسر والأفراد.

طرق التحقق من صحة البيانات

تتمثل إحدى طرق التحقق المهمة في تحليل تماسك نتائج المسح مع مصادر البيانات الأخرى، ولا سيما بالنسبة للمسوحات الأسرية، ويمكن استخدام ما يلي للتحقق من تماسك النتائج:

- بيانات السكان والمساكن
  - إجمالي عدد الأسر في أبو ظبي في سنة تقويمية
- مصادر البيانات الإدارية (مثل السجل السكاني)

طرق التعويض

يرجى مراجعة القسم 5.4 أعلاه لوصف العمليات المعنية بتعويض القيم المفقودة.

احتساب الأوزان

يجب احتساب وزن الوحدات الفردية للعينة لتمثيل المجتمع بأكمله، حساب الاوزان عملية مشتقة من تصميم العينة (التقسيم الطبقي والمعاينة متعددة المراحل والمعاينة العنقودية وما إلى ذلك). وبالتالي، لا يمكن تحديده في هذه المرحلة، وعلى الرغم من ذلك، يوصى بأن يتبع تصميم العينة تصميم المسوح الأسرية الأخرى (على سبيل المثال، مسح القوى العاملة.

تجدر الإشارة إلى أنّه، في حال شكّلت الأسرة وحدة المعاينة النهائية (أي في حال تمّ مقابلة جميع أفراد الأسرة وفقًا للتوصيات الدولية)، يتمّ احتساب أوزان العينات على مستوى الأسرة، في حال تمّ إجراء عملية معاينة لاحقة (أي في حال تمّ اختيار شخص واحد للإجابة من كلّ أسرة)، يتعيّن احتساب أوزان العينات على المستوى الاسرة.

تجدر الإشارة إلى أنّ حساب الاوزان يكون أكثر تعقيدًا في حال تضمن تصميم المعاينة التقسيم الطبقي و/أو متعدد المراحل و/أو المعاينة العنقودية وذلك على النحو الموصي به في الوثائق المنهجية المُعدّة لكل مسح.

عادةً ما تُحتسب المؤشرات على أنّها نسب الأرقام المجمعة بعد المعاينة.

3.2 مسوح المنشآت الاقتصادية

يصف هذا القسم خصائص طرق المعالجة والتحقق من الصحة ومراقبة الجودة المذكورة أعلاه، والمطبقة في مسوح المنشآت.

في حالة المسوح الاقتصادية، غالبًا ما تتضمن استمارة المسح قسمًا لجمع البيانات على مستوى المنشأة الرئيسية، وقسمًا لجمع البيانات على مستوى كل منشأة أو وحدة محلية، أو على مستوى وحدة نوع النشاط. يُترجم هذا إلى قاعدة بيانات والتي تحتوي على البيانات على مستويات مختلفة، مرتبطة من خلال معرفات فريدة للمنشات والمؤسسات

طرق التحقق من صحة البيانات

تغطي عملية تدقيق البيانات العمليات الفرعية المُشار إليها بلفظ "التدقيق الجزئي" و"التدقيق الكلي" (كما يمكن الإشارة إليها بلفظ تدقيق "المدخلات" وتدقيق "المخرجات"):

- التدقيق الجزئي: تشير هذه العملية إلى التحقق من الصحة والتعديلات المطبقة على بيانات منشأة معنية، وتتضمن معالجة البيانات الناقصة أو المفقودة، واكتشاف الإجابات غير المتسقة داخليا مع الأسئلة الأخرى ومعالجتها.

تشير عملية التدقيق الكلي إلى التحقق من الصحة والتعديلات المرتبطة بمجموعات البيانات الكاملة، عن طريق تحليل المجاميع، وتهدف إلى التحقق من توافق بعض التقديرات بشكل مشترك مع المعلومات المعروفة الأخرى، وتتضمن عملية التدقيق الكلي المعقد، إعادة ضبط أوزان العينة وفقًا لأخطاء الإطار التي تمّ اكتشافها أثناء إجراء المسح.

وللحدّ من هذه الأخطاء، يكون من الضروري تطبيق التقنيات التي تحسّن فعالية أدوات جمع البيانات وإجراءات الجمع، بالإضافة إلى ذلك، يتعيّن استخدام تقنيات قوية لتدقيق البيانات، وذلك لتحويل البيانات الأولية المقدمة من المستجيبين، إلى بيانات صحيحة ومتماسكة ("نظيفة")، بحيث يكون من الممكن استخدامها لإنتاج احصاءات مجمعة.

معالجة حالات عدم الاتساق الداخلي والأخطاء

تتضمن عملية تدقيق البيانات التحقق من البيانات الأصلية ومعالجتها في بعض الأحيان، لأنه قد ينجم عن هذه العمليات أخطاء تؤثر على مجموع البيانات، وبالرغم من أهمية عملية تدقيق البيانات، فمن الضروري إنشاء الممارسات التي تحدّ من حدوث البيانات غير الكاملة أو غير المتسقة، بهدف الحدّ من تأثير تدقيق البيانات إلى أدنى حدّ، كما ستقوم ضوابط الجودة المضمنة بالفعل في وسائل جمع البيانات أو في مرحلة إدخال البيانات، بتحسين نوعية البيانات الأولية والتخفيف من مهمة معالجة البيانات.

ان اختيار أداة الجمع له تأثير مباشر على جودة البيانات، بحيث يمكن توقع أن تؤدي كل من المقابلات الشخصية بمساعدة الكمبيوتر (CAPI) والمقابلات الهاتفية بمساعدة الكمبيوتر (CATI) إلى تحسين جودة بيانات الإدخال، لأنها توفر ضوابط تلقائية لاكتشاف أخطاء الاستجابة.

يتكون التحكم في الصلاحية لعنصر بيانات فردي من؛ التحقق مما إذا كانت الإجابة تنتمي إلى مجموعة محددة مسبقًا (أو نطاق) من الاستجابات الصالحة، من أجل التحقق من صحة الأسئلة، فمن الضروري التحقق منها مقابل تلك الإجابات الصحيحة المحددة، وللتحقق من التناسق الداخلي للاستمارة، فمن الضروري إنشاء وتطبيق القواعد التي تحدد العلاقات بين الأسئلة، بحيث تقيد بعض إجابات القيم الصالحة التي يمكن أن تقبلها الأسئلة الأخرى، وكما يمكن تطبيق عمليات التحقق الحسابية (على سبيل المثال، مجموع النسب المئوية للتوزيعات التي تصل إلى 100)، أثناء إدخال البيانات، أو في وقت لاحق أثناء عملية وضع الدفعات اللاحقة عبر مجموعة من السجلات.

معالجة الوحدات المصنّفة بشكل خاطئ

هناك مشكلة متكرر تؤثر على جودة إحصاءات المنشآت، تتمثل بإدراج بعض المنشآت في البداية في الطبقة الخطأ ضمن إطار المجتمع الذي يتمّ اختيار العينة منه، وتزداد احتمالية حدوث ذلك في الحالات التي تكون جودة الإطار (وسجل المنشأة الأساسي) منخفضة، وعادةً ما تحتوي السجلات التي تحتفظ بها مكاتب الإحصاء الوطنية على معلومات عن حجم المنشأة (فيما يتعلق بعدد العمال و/أو معدل الدوران)، بالإضافة إلى النشاط الاقتصادي والموقع (بالاستناد إلى عنوان المنشأة)، ومن الممكن إدخال الوحدات المصنّفة بشكل خاطئ بصفتها صحيحة، وبما انها مصنفة بشكل خاطئ فهي لن تظهر ضمن الإطار، أو تظهر في الطبقة الخاطئة، ففي الحالة الأولى؛ عند استبعاد المنشأة التي تمّ إجراء مسح لها من العينة، بسبب عدم أهليتها للخضوع للمسح، سيحدّ ذلك من حجم العينة الفعال ما لم يتمّ إعداد قائمة بالاحتياطي، وهنا يجب عدم النظر الى استبعاد الوحدات المصنّفة بشكل خاطئ إلّا في الحالات التي يكون فيها معدّل التصنيف الخاطئ صغيرًا، أمّا في الحالة الثانية، فتكون العينة مؤهلة للمسح، إلّا أنّه تم إدراجها في الطبقة الخاطئة أو حذفها من الإطار. فعلى سبيل المثال، قد تعلن منشأة مختارة في فئة الحجم (الطبقة) من 10 إلى 20 موظفًا، أن لديها 8 موظفين فقط، وهنا يكون الحلّ الفني في إعادة احتساب أوزان العينة، وبالتالي إنتاج تقدير جديد لحجم الطبقات، وتصحيح الأوزان وفقا لذلك، ومن الواضح أنّ إنشاء وصيانة سجل منشآت حديث، يمكن من خلاله رسم إطار موثوق به، هو أمر بالغ الأهمية.

وحيث أنه من غير العملي إعادة الاتصال بالمستجيبين، من الممكن تقدير (التعويض عن) البيانات المفقودة. فإذا كانت إعادة الاتصال بالعمل الذي تمت مقابلته خارج النطاق لأسباب تتعلق بالتكلفة أو الوقت، فيمكن معالجة عدم الاستجابة للعنصر من خلال الأساليب الرياضية لتعويض البيانات، وتتكون هذه العملية من تخصيص قيمة معقولة لسؤال لم تقدم الوحدة المختارة إجابة له، أو لسؤال تتعارض إجابته منطقيًا أو حسابيًا مع الإجابات في بقية الاستمارة، وفي حالة عدم اتساق الإجابة لعدد من الأسئلة، فقد يصعب تحديد الإجابة الخاطئة. وعادة ما يتمّ إنشاء تسلسل هرمي بين الأسئلة، أو بين مجموعات الأسئلة، بحيث تعتبر الإجابات على بعضها "الصفة الغالبة."

تتمثل إحدى طرق التعويض المستخدمة في المسوح المتكررة للمنشآت في تعيين الإجابة التي تقدمها الوحدة نفسها في مسح تمّ إجراؤه مسبقًا (التعويض التاريخي)، ويمكن تطبيق المبدأ نفسه في حالات عدم استجابة الوحدة، ويتم تطبيق هذه التقنية بشكل متكرر في حالات المنشآت الكبيرة التي يكون من المرجح أن تخضع للمسوح المتتالية. يرجى الرجوع إلى القسم الرابع أدناه لوصف عمليات تعويض القيم المفقودة.

طرق التعويض

يرجى مراجعة القسم 5.4 أعلاه لوصف العمليات المعنية بتعويض القيم المفقودة.

حساب الأوزان

من المعروف أن وزن وحدة العينة هو معكوس احتمال اختيار الوحدة المعنية من العينة. تستخدم الأوزان لتمثّل المجتمع المستهدف لأغراض معايرة البيانات (normalization purpose). ضمن تصميم العينات الطبقية، يتم حساب أوزان العينات للمنشآت داخل كل طبقة بشكل مستقل. وزن المنشأة الاقتصادية ضمن طبقة معينة يساوي ناتج قسمة عدد المنشآت الاقتصادية في الإطار على عدد المنشآت الاقتصادية المستجيبة لبيانات المسح.

أما بالنسبة لطبقة المنشآت الكبيرة، وخاصة الطبقة الجزئية الأولى، فإن وزن المنشأة الاقتصادية يجب أن يكون مساوياً لـ 1، أي أنها تمثل نفسها فقط، على اعتبار أنه قد تم أخذها باحتمال 100%. أما باقي المنشآت ضمن الطبقة الجزئية الثانية فيحسب وزن كل منها بنفس الالية السابقة.

5. المرجع

UNECE Statistics Wikis - Generic Statistical Data Editing Models, https://statswiki.unece.org/display/sde/GSDEMs. (Downloaded 6th July 2023)

مـركـز الإحـصـاء
STATISTICS CENTRE

الرؤية: ببياناتنا نمضي نحو غدٍ أفضل
Vision: Driven by data for a better tomorrow

www.scad.gov.ae
adstatistics