	METODOLOGÍA PARA LA INVESTIGACIÓN DE ANALÍTICA DE DATOS		
	Proceso: Gestión del Conocimiento	Versión: 1	Código: MT1.GC
			Fecha Aprobación: 21/12/2020

TABLA DE CONTENIDO

El presente documento está conformado por 4 secciones posteriores a la introducción. Después de una breve presentación, se exponen el objetivo general y los objetivos específicos de la metodología para la investigación de analítica de datos. A continuación, se exponen el objetivo general y los objetivos específicos de esta metodología. En segundo lugar, se presenta el marco conceptual, seguido por las referencias normativas. Por último, se expone la descripción de la metodología.

0. Introducción
1. Objetivo general y específicos
2. Marco conceptual
3. Referencias normativas
4. Descripción de la metodología

INTRODUCCIÓN

Según el Decreto 588 de 2017, dentro de las funciones de la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición (en adelante CEV) se encuentra “Investigar todos los componentes de su mandato, a través de las metodologías y formas de recolección y análisis de información necesarias para tal efecto, considerando las generalmente aceptadas por las ciencias sociales, con un enfoque de género, y teniendo en cuenta los anteriores esfuerzos de construcción de la verdad, incluyendo como insumo básico, entre otros, los informes de la Comisión Histórica del Conflicto y sus Víctimas”. Asimismo, en el artículo 14 se establece que la CEV debe “(tomar las) medidas necesarias para garantizar la mayor objetividad e imparcialidad”. Con estos antecedentes, la CEV creó la “Guía metodológica para la verificación, contrastación y profundización de la información”, en la que se “desarrolla un procedimiento de valoración con criterios específicos para la verificación, contraste y profundización de los hallazgos de la Comisión de la Verdad relacionados con hechos, casos, patrones y contextos explicativos”.

Dado que la analítica de datos, según el procedimiento con fecha de aprobación 06/03/2020 y disponible en el mapa de procesos, “es un procedimiento orientado al óptimo aprovechamiento de los grandes volúmenes de información misional para el entendimiento de las dinámicas, los patrones y las alternativas al conflicto en contexto, abordando desde esta perspectiva las principales preguntas e hipótesis de investigación”, es necesario que cumpla con una metodología definida. Esto para garantizar la transparencia, auditabilidad y replicabilidad de los ejercicios cuantitativos realizados por el equipo.


En el presente documento se expone entonces cómo se debe regir la investigación de analítica de datos para cumplir con los principios mencionados previamente y garantizar la rigurosidad e imparcialidad dentro de los análisis cuantitativos realizados por la CEV.

1. OBJETIVO GENERAL Y ESPECÍFICO:

Objetivo general:

Piensa en el medio ambiente, antes de imprimir este documento.

Cualquier copia impresa de este documento se considera como COPIA NO CONTROLADA

	METODOLOGÍA PARA LA INVESTIGACIÓN DE ANALÍTICA DE DATOS		
	Proceso: Gestión del Conocimiento	Versión: 1	Código: MT1.GC
			Fecha Aprobación: 21/12/2020

- Establecer una metodología para la investigación cuantitativa que permita cumplir con los principios de transparencia, replicabilidad y auditabilidad para garantizar la rigurosidad e imparcialidad de los análisis cuantitativos realizados por la CEV.

Objetivos específicos:

- Definir los requisitos mínimos con los que debe cumplir la investigación cuantitativa
- Establecer mecanismos para garantizar transparencia, replicabilidad y auditabilidad de la investigación cuantitativa
- Generar espacios para asegurar la imparcialidad y rigurosidad de la investigación cuantitativa

2. MARCO CONCEPTUAL


La investigación cuantitativa reduce una serie de variables a un conjunto pequeño y controlado para probar una hipótesis, que acompañado de métodos o instrumentos de validación permiten darle una interpretación a los datos (Cresweel, J, 1994). Por medio del uso de estadística es entonces posible validar o falsear hipótesis, así como predecir variables. Para esto, se deben seleccionar métodos estadísticos según criterios como:

“a) el nivel de medición de las variables (nominal, ordinal, intervalar o de razón), b) la distribución teórica de los datos (normal o no), que lleva al empleo de estadística paramétrica o no paramétrica, c) el número de grupos de comparación: dos o más, d) la existencia de un efecto de interacción entre las variables independientes o no, sobre la variable dependiente, e) la variable dependiente es unidimensional o multidimensional, en el segundo caso se requerirá de modelos estadísticos multivariados, f) predecir una variable a partir de otras, mediante diversos tipos de regresiones múltiples, g) si el objetivo es desarrollar un modelo explicativo, se requiere el modelamiento de ecuaciones estructurales, por ejemplo” (Reidl-Martínez, L., 2012)).

Además de estos criterios, es fundamental la validación de supuestos y su solución. Sin embargo, un mal uso de los datos puede llevar a conclusiones equivocadas. Por lo tanto, es necesario que la investigación realizada por la CEV cumpla con los principios de transparencia, auditabilidad y replicabilidad. Para esto, la organización Human Rights Data Analysis Group (en adelante HRDAG) creó un “flujo de trabajo cuántico”, que sigue una estructura de trabajo de pequeñas piezas unidas para reducir al máximo los errores y dar mayor flexibilidad. En este flujo de trabajo, se crean tareas autocontenidas y autodocumentadas, que evitan que la información pueda ser eliminada y facilita la revisión, corrección y acceso a la información (Ball, P., 2016).

3. REFERENCIAS NORMATIVAS

Decreto 588 de 2017 “Por el cual se organiza la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la no Repetición”.

	METODOLOGÍA PARA LA INVESTIGACIÓN DE ANALÍTICA DE DATOS		
	Proceso: Gestión del Conocimiento	Versión: 1	Código: MT1.GC
			Fecha Aprobación: 21/12/2020

4. DESCRIPCIÓN DE LA METODOLOGÍA

Dentro de esta metodología participan los siguientes actores con sus respectivos roles:

Profesionales de la Dirección de Conocimiento, Dirección de Territorios, Dirección de Pueblos Étnicos o Dirección de Diálogo Social y Profesionales de la Coordinación del SIM:

- definir las hipótesis a ser validadas por medio de métodos cuantitativos.
- aplicar métodos cuantitativos para responder a preguntas de investigación que contribuyan al esclarecimiento.

Director(a) de Conocimiento, Coordinador SIM, Profesionales de la Dirección de Conocimiento, Dirección de Territorios, Dirección de Pueblos Étnicos o Dirección de Diálogo Social, Profesionales de la Coordinación del SIM:

- asistir a los seminarios periódicos de revisión de pares para garantizar la imparcialidad y rigurosidad de las investigaciones cuantitativas.


En primer lugar, es necesario definir las hipótesis que se van a validar por medio de métodos cuantitativos. Para esto, se debe delimitar al máximo la pregunta de investigación, incluyendo la población, periodo, lugar u objeto de estudio. Con la pregunta definida y delimitada es entonces posible entrar a estudiar con qué información estructurada cuenta la CEV para responderla, para lo que se debe explorar el catálogo de microdatos - CKAN.

Al tener claridad sobre la información disponible, es posible entonces establecer los posibles métodos cuantitativos a utilizar para responder la pregunta de investigación. Con la definición de las bases de datos y posibles métodos, se debe generar un cronograma de trabajo que permita tener resultados de manera oportuna para impulsar la investigación en la Comisión.

En aras de garantizar la imparcialidad y rigurosidad del proyectos, se hace necesaria una revisión de pares, que indaguen sobre la selección de la base de datos, variables y metodología. De esta manera, la discusión colectiva permite la implementación del proyecto más riguroso posible e imparcial posible.

Después de definida la metodología, el investigador inicia el trabajo de programación. Dado que en este punto podría haber errores humanos que lleven a conclusiones equivocadas, el profesional debe seguir el flujo de trabajo definido por HRDAG con tareas autocontenidas y autodocumentadas, que permitan una revisión sencilla del código por los pares. Para evitar que haya pérdida de información, así como promover el trabajo colaborativo y dar acceso a todos los especialistas en analítica de datos al código, se debe crear un proyecto en GitLab. Dentro del repositorio del equipo de analítica del SIM se genera entonces un proyecto propio para la investigación. Todos los scripts desarrollados en el proyecto deben estar en código abierto.

El proyecto debe tener un archivo README, que explique la pregunta de investigación y el flujo de trabajo a seguir. El flujo de trabajo debe mostrar las tareas de manera numerada y estas deben ser autocontenidas y autoexplicativas. La tarea inicial suele ser “importar”, donde se traen los datos del archivo .csv o .xls de CKAN para transformarlos al formato correspondiente de R o Python. A continuación, la tarea suele ser “estandarizar” para, como su nombre indica, estandarizar las bases de

	METODOLOGÍA PARA LA INVESTIGACIÓN DE ANALÍTICA DE DATOS		
	Proceso: Gestión del Conocimiento	Versión: 1	Código: MT1.GC
			Fecha Aprobación: 21/12/2020

datos o “limpiar” para homogeneizar, por ejemplo, los nombres de los municipios. Otra tarea puede ser “modelar” para correr el modelo, y así sucesivamente. El profesional tiene la libertad de definir la cantidad de tareas que sean necesarias, siempre y cuando en una misma tarea no se esté incluyendo el trabajo de dos.

Cada una de las tareas debe tener su propia carpeta y dentro de la misma debe haber al menos otras dos carpetas: src, en donde se almacenan los scripts, y output, donde se almacenan los resultados de la tarea. También es posible tener una carpeta de “input”, en la que se almacenan los datos de entrada a ser ejecutados por el script. Finalmente, debe haber un Makefile, que permita automatizar la tarea.

Independientemente de la cantidad de tareas que el profesional haya especificado, el proyecto siempre debe incluir una última tarea de “escribir” en la que, como su nombre indica, se escriban los resultados del proyecto. Esta tarea debe tener como input los output que fueron generados en otras tareas (como gráficas, mapas, entre otras) y yaml files con los resultados que serán llamados en el informe. Dentro de la carpeta src debe haber un archivo .Rmd en el que se realice la escritura. Por último, en la carpeta output debe estar el informe de resultados en .PDF.

Una vez se tenga el informe de resultados, se debe tener un espacio de seminario con pares. De esta forma, el investigador recibe retroalimentación sobre la aplicación de la metodología, así como de la interpretación de los resultados. En caso de ser necesario hacer ajustes, debe implementarlos en el informe de resultados.


Tan pronto el informe de resultados esté terminado, el profesional debe dejar explícito en el README del proyecto “Este proyecto fue terminado el día XX de XX de 202X”. Asimismo, debe difundir el reporte con los equipos interesados.

Estos pasos permiten garantizar la rigurosidad e imparcialidad de la investigación cuantitativa. Además, almacenar la información en el Git permite cumplir con los principios de transparencia, replicabilidad y auditabilidad. Transparencia, porque cualquier persona puede acceder a los códigos y estudiarlos, pues al ser en código abierto no es necesaria una licencia. Replicabilidad, ya que el Makefile debe poder ejecutarse desde cualquier computador, lo que garantiza que cualquier individuo pueda obtener exactamente los mismos resultados que el profesional ha generado previamente. Auditabilidad, debido a que al tener pequeñas tareas es posible identificar errores de una manera más inmediata que enfrentándose a largos scripts.

5. DOCUMENTOS DE REFERENCIA

Ball, P. 2016. The Task Is A Quantum Of Workflow. HRDAG. Disponible en: <https://hrdag.org/2016/06/14/the-task-is-a-quantum-of-workflow/>. Recuperado el 30 de noviembre de 2020

Comisión de la Verdad. 2020. Guía metodológica para la verificación, contrastación y profundización de la información”, en la que se “desarrolla un procedimiento de valoración con criterios específicos para la verificación, contraste y profundización de los hallazgos de la Comisión de la Verdad relacionados con hechos, casos, patrones y contextos explicativos.

	METODOLOGÍA PARA LA INVESTIGACIÓN DE ANALÍTICA DE DATOS		
	Proceso: Gestión del Conocimiento	Versión: 1	Código: MT1.GC
			Fecha Aprobación: 21/12/2020

Comisión de la Verdad. 2020. Procedimiento de analítica de datos. Disponible en: <https://comisiondelaverdad.co/la-comision/mapa-de-procesos/gestion-del-conocimiento/1-procedimientos/p5-gc-analitica-de-datos>. Recuperado el 30 de noviembre de 2020.

Congreso de la República de Colombia. 2017. Decreto 588 de 2017

Creswell, J. 1994. Research design. 4th edition

Reidl-Martínez, L. 2012. Marco conceptual en el proceso de investigación. Metodología de la Investigación en Educación Médica. 2012;1(3):146-151

VALIDACIÓN	NOMBRE	CARGO	FECHA
Elaboró	Valentina Rozo	Analista Coordinación del Sistema de Información Misional	21/12/2020
Revisó	Andrés Medina	Coordinador Sistema de Información Misional SIM	21/12/2020
Aprobó	Diana Britto	Directora de Conocimiento	21/12/2020