

Informe de la fase de procesamiento

 Cuentas Satélite de Educación
(CSE) 2023

Diciembre, 2024



Tabla de contenido

1.	Introducción.....	3
2.	Objetivo	3
3.	Desarrollo	3
3.1.	Críticar e integrar las bases de datos	4
3.1.1.	Ejecutar la crítica de la información recolectada (5.1.1)	4
3.1.2.	Ejecutar la integración/sincronización de la base de datos desde las distintas fuentes planificadas (5.1.3)	4
3.1.3.	Aplicar los controles de confidencialidad en el procesamiento (5.1.4)	5
3.2.	Clasificar y/o codificar	5
3.2.1.	Clasificar y codificar la base de datos (5.2.1)	5
3.3.	Validar e imputar	5
3.3.1.	Validar la base de datos (5.3.1)	6
3.3.1.	Aplicar métodos de imputación a la base de datos (5.3.2)	8
3.4.	Derivar nuevas variables y unidades	8
3.4.1.	Calcular variables derivadas (5.4.2)	8
3.5.	Finalizar los archivos de datos.....	8
3.5.1.	Compilar las bases de datos derivadas del proceso (5.7.1)	8
3.5.2.	Compilar los productos estadísticos de los procesos anteriores, plan de tabulados, fichas de indicadores, informes, entre otros documentos finales	9
4.	Conclusiones	9

Índice de tablas

Tabla 1.	Variables definidas para las bases de datos intermedias.....	7
-----------------	--	---



1. Introducción

El Instituto Nacional de Estadísticas y Censos (INEC) dentro de su programación estadística plurianual y en el marco del Programa Nacional de Estadística elabora las Cuentas Satélite de Educación (CSE), como una herramienta de cuantificación económica de los flujos de oferta y demanda de servicios de educación en el país. Es así que la Dirección de Estadísticas Económicas (DECON), a través de la Gestión de Análisis de Síntesis (GASIN), planificó en el año 2024 la construcción de las Cuentas Satélite de Educación (CSE) periodo 2021-2023 que se publican en octubre del presente año.

Es importante destacar que las actividades de las CSE se alinean a las fases del Modelo de Producción Estadística (MPE), facilitando así los procesos de ejecución y seguimiento de esta operación estadística según las buenas prácticas estadísticas. Las fases del MPE son las siguientes:

- Planificación
- Diseño
- Construcción
- Recolección
- **Procesamiento**
- Análisis
- Difusión
- Evaluación

La fase de procesamiento comprende un conjunto de actividades de depuración de datos, integración de información, preparación de bases de datos intermedias, cálculo de variables derivadas y preparación de información para la etapa de análisis y posterior difusión. Como se ha mencionado, es una etapa en la que se reprocesa la información varias veces hasta obtener bases de datos intermedias depuradas y con información de calidad para la generación de resultados en la fase de análisis.

De este modo, el presente documento recoge las actividades realizadas durante la fase de procesamiento de las CSE, como: la codificación aplicada a la cobertura de las instituciones del sector de la educación y métodos de crítica, validación e integración aplicada a la base de datos que se utilizan para construir las CSE.

2. Objetivo

Documentar las actividades realizadas en la fase de procesamiento como parte de la generación de las Cuentas Satélite de Educación 2021-2023.

3. Desarrollo

La CSE es una operación estadística de síntesis que utiliza una gama de información de operaciones estadísticas de base, registros administrativos, registros estadísticos y operaciones estadísticas de síntesis, generadas por instituciones como el Banco Central del Ecuador (BCE), el Ministerio de Educación (Mineduc), el Ministerio de



Inclusión Económica y Social (MIES), la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (Senescyt), entre otras.

El mayor esfuerzo de una operación estadística de síntesis no se concentra en la recolección de información, como sucede en un proceso típico de levantamiento de información a través de una encuesta o un censo, sino que se concentra en mayor proporción en las etapas de tratamiento e integración de bases de datos o registros de datos primarios y en la generación de bases de datos intermedias obtenidas a partir de la información previamente captada.

Los procesos y actividades de esta fase se ejecutan de acuerdo con las buenas prácticas estadísticas que buscan generar datos con características de precisión, confiabilidad, coherencia y comparabilidad. En esta línea, a continuación se describen los procesos establecidos conforme al nivel 3 del MPE implementadas para generar las CSE por el equipo técnico de la Gestión de Análisis de Síntesis (GASIN), durante la fase de procesamiento.

1. Criticar e integrar la base de datos
2. Clasificar y/o codificar
3. Validar e imputar
4. Derivar nuevas variables y unidades
5. Ajustar los factores de expansión
6. Verificar la representatividad
7. Finalizar los archivos de datos

3.1. Criticar e integrar las bases de datos

En este proceso, se revisa la información recolectada, es decir, se analiza que cada base de datos recolectada contenga las variables necesarias y los datos guarden coherencia a lo largo de la serie (niveles dentro de un rango coherente) y entre otros criterios a fin de que no existan errores en la información recolectada. Este proceso se constituye como un primer filtro para evaluar la calidad de la información antes de su integración.

Posteriormente, se integran las bases de datos, unificando todas las bases criticadas de las diferentes instituciones proveedoras de información. Tras la integración, en función de los requisitos de protección de datos, los datos deben ser anonimizados (limitando en lo posible el uso de variables de identificación de los microdatos). Las actividades realizadas en este proceso son las siguientes:

3.1.1. Ejecutar la crítica de la información recolectada (5.1.1)

Esta actividad se ejecuta como parte del proceso 5.3.2 "Aplicar métodos de imputación a la base de datos".

3.1.2. Ejecutar la integración/sincronización de la base de datos desde las distintas fuentes planificadas (5.1.3)

Esta actividad se ejecuta como parte del proceso del punto 5.3.2 "Aplicar métodos de imputación a la base de datos".



3.1.3. Aplicar los controles de confidencialidad en el procesamiento (5.1.4)

El verificable de esta actividad es el mismo que se ha definido para el punto 2.1.3 "Identificar y diseñar procedimientos que rigen el control y acceso a la información confidencial en todo el proceso de producción" de la fase de Diseño y se encuentra en la siguiente ubicación:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\2_Dis\2.1_Dis_prod\2.1.3_Proc_info_confid

Archivos: 1_Contr_confidenc_CSE_2023 y 2_Reg_mane_info_sensi_2023

3.2. Clasificar y/o codificar

Las CSE clasifican y codifican las bases de datos en función de los parámetros determinados en el diseño metodológico y nomenclaturas. Esta clasificación permite estructurar tabulados e indicadores en concordancia con el marco central de Cuentas Nacionales y además favorece la comparabilidad con clasificadores internacionales.

Las principales clasificaciones comprenden:

- Clasificar y codificar las nomenclaturas y correspondencias de las Cuentas Satélite de Educación.
- Clasificar y codificar las partidas presupuestarias de ingresos y gastos del sector público y privado de acuerdo a los códigos de cuentas nacionales.
- Clasificar y codificar los establecimientos según la Clasificación Internacional de Industrias Uniforme CIIU.

Una vez generadas estas relaciones, se genera la base codificada y clasificada de acuerdo a las nomenclaturas mencionadas.

3.2.1. Clasificar y codificar la base de datos (5.2.1)

Esta actividad se ejecuta como parte del punto 5.3.2 "Aplicar métodos de imputación a la base de datos".

3.3. Validar e imputar

El proceso de validación en las bases clasificadas y codificadas hace referencia a los siguientes pasos:

- Se observa la tendencia de la serie para detectar datos atípicos utilizando cuadros y gráficos estadísticos, los cuales permiten notar la tendencia de la serie del tiempo, índices simples, tasas de participación y variaciones anuales.
- Una vez que se han llevado a cabo estas actividades y se ha logrado identificar posibles casos atípicos dentro de la serie, los mismos son



analizados con el fin de determinar las razones del fenómeno; los cuales a través de procesos estadísticos y analíticos pueden ser corregidos o justificados.

- Es importante también destacar que los resultados se comparan con otras variables del contexto económico y social del país, con la finalidad de verificar la confiabilidad y consistencia de los resultados obtenidos.

3.3.1. Validar la base de datos (5.3.1)

La validación implica analizar y revisar exhaustivamente cada base de datos utilizada como fuente de información de las CSE. Por ello, en este punto se analizan las bases de datos de forma individual según la información recolectada.

Es importante señalar que las bases de datos recolectadas no tienen la misma estructura ni número de variables, por ende, no se puede unificar la información antes de prepararla y analizarla.

Las bases de datos que se utilizan y analizan dentro del proceso de construcción de las CSE son:

Sector privado

- Base de datos de Sociedades no financieras e ISFLSH
- Base de datos de Cuasi sociedades:
- Base de datos de Hogares
- Base de datos de Conexos
- Base de datos de Universidad Andina Simón Bolívar
- Base de datos de FLACSO

Sector público

- Base de datos del Gobierno Central (Ministerio de Economía y Finanzas)
- Base de datos del Ministerio de Educación
- Base de datos de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación.
- Base de datos del Ministerio de Inclusión Económica y Social
- Consejo de Educación Superior.
- Base de datos de Gobiernos locales
- Base de datos de Secretaría de Educación del Distrito Metropolitano de Quito
- Base de datos de Patronato San José
- Base de datos del Ministerio de Transportes y Obras Públicas

Las bases de datos que se analizan en el sector privado y público poseen información sensible (microdato), por tanto, además de procesar la información de manera consistente y precisa, esta debe estar sujeta a las normas de confidencialidad de la información. Cada base de datos al final del procesamiento debe estar conformada por un conjunto de 26 variables homologadas, las cuales posteriormente integran la base de datos unificada final.



Las variables que conforman las bases de datos individuales y la base de datos unificada se detallan a continuación:

Tabla 1. Variables definidas para las bases de datos intermedias

Nº	Variable	Descripción
1	ejercicio	Año o ejercicio de la información
2	tipo	Tipo de presupuesto: 1 = ingreso 2 = gasto
3	id_registro	Código de registros para tratamientos especiales, se utiliza para las matrices de distribución.
4	id_cs	Código de la entidad institucional – establecimiento o institución
5	codigo_N1	Código a nivel 1 de las nomenclaturas de las CSE
6	codigo_N6	Código a nivel 6 de las nomenclaturas de las CSE 2020-2023
7	descr_codigo_N6	Descripción del código a nivel 6 de las nomenclaturas de las CSE 2020-23
8	codigo_N6_ant	Código a nivel 6 de las nomenclaturas de las CSE 2018-2019
9	descr_codigo_N6_ant	Descripción del código a nivel 6 de las nomenclaturas de las CSE 2018-19
10	cod_programa	Código del programa
11	descr_programa	Descripción del programa
12	cod_proyecto	Código del proyecto
13	descr_proyecto	Descripción del proyecto
14	cod_actividad	Código de la actividad
15	descr_actividad	Descripción de la actividad
16	cod_provincia	Código de la provincia
17	descr_provincia	Descripción de la provincia
18	cod_canton	Código de cantón
19	descr_canton	Descripción del cantón
20	part	Código de la partida presupuestaria a 6 dígitos
21	descr_part	Descripción de la partida presupuestaria a 6 dígitos
22	cod_CN	Código según Cuentas Nacionales
23	descr_cod_CN	Descripción del código de Cuentas Nacionales
24	devengado	Valor presupuestario devengado
25	fuentes	Descripción de la fuentes que emite la información
26	observaciones	Nota aclaratoria de un dato o tratamiento realizado a la base individual

Fuente: INEC, CSE

Otras bases de datos

- Base de datos de matrices de distribución EGB y desarrollo infantil
- Base de datos del matrices de 3 y 4 nivel
- Bases de datos para desagregación territorial
- Base de datos de deflatores
- Bases de datos de equilibrios conexos
- Bases de programas del Mineduc

Las bases antes mencionadas se utilizan para la generación de las bases de datos de síntesis, por lo que constituyen un insumo fundamental para la generación de las CSE. Las bases de datos se encuentran en la siguiente ubicación:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\5_Proc\5.3_Valid_input\5.3.1_Valid_bas\3_Otros

3.3.1. Aplicar métodos de imputación a la base de datos (5.3.2)

En esta sección, se elaboran los informes referentes a las bases de datos construidas para el sector público, privado y otros. Los informes generados son el soporte y respaldo de la implementación de procesos para la generación de las CSE.

Los documentos verificables de este punto se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\5_Proc\5.3_Valid_imput\5.3.2_imput_dat

3.4. Derivar nuevas variables y unidades

En este proceso se generan variables propias de las cuentas satélites que nos permitirán obtener tablas y cuadros de síntesis económica.

A manera de ejemplo, en las CSE se genera la base de datos intermedia denominada "economía total" que recoge información procesada para la generación de agregados económicos como los contenidos en las cuentas corrientes. Concretamente, en este punto se generan variables según Cuentas Nacionales de empleos y recursos para cada unidad institucional de las CSE.

3.4.1. Calcular variables derivadas (5.4.2)

En esta actividad se detallan y generan todos los demás insumos necesarios para la generación de las bases intermedias como: scripts usados para el procesamiento de las bases de datos de síntesis y tabulados y bases de trabajo. Los archivos se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\5_Proc\5.4_Deriv_variab\5.4.2_Cal_variab_deriv

3.5. Finalizar los archivos de datos

Una vez finalizado el procesamiento de las bases individuales (mencionadas en el punto 3.3.1) y a partir de las 26 variables homologadas en cada base individual, estas se integran en un solo bloque de información denominada Base Unificada de las CSE.

3.5.1. Compilar las bases de datos derivadas del proceso (5.7.1)

La base de datos unificada que recopila todas las bases de datos del sector público y privado, es el insumo más importante dentro de la fase de procesamiento de las CSE. Pues de aquí, se derivan y construyen las bases de datos de síntesis intermedia que durante la fase de Análisis permitirán construir los tabulados e indicadores económicos.



Los documentos verificables incluyendo la base unificada se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\5_Proc\5.7_Finali_archiv_dat\5.7.1_Compil_bas_dat\1_BDUCSE_2021_23

3.5.2. Compilar los productos estadísticos de los procesos anteriores, plan de tabulados, fichas de indicadores, informes, entre otros documentos finales

En este punto, se procede a realizar el cálculo y procesamiento de los tabulados. Para ello, se construyen scripts en el programa estadístico R y se utilizan plantillas específicas para cada tipo de tabulado como:

- Cuentas corrientes
- Cuadros de financiamiento y erogaciones
- Tablas de oferta y utilización
- Cuadros de producción, consumo intermedio y valor agregado bruto
- Indicadores económicos
- Indicadores de financiamiento y erogaciones
- Indicadores físicos
- Indicadores territoriales

Finalmente, se llevan a cabo revisiones y correcciones, que pueden implicar reprocesos de información, según el contenido de cada tabulado, hasta obtener las estadísticas de síntesis e indicadores económicos finales que pasarán a la fase de Difusión.

Los documentos verificables de este punto se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

Ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2024\CSE_2021_23\5_Proc\5.7_Finali_archiv_dat\5.7.2_Compil_prod_ant

4. Conclusiones

- En la fase de procesamiento se llevaron a cabo actividades de crítica, codificación y validación de los datos recolectados, así como la generación de los tabulados e indicadores de las CSE 2021-2023.
- Las actividades de la fase de procesamiento, alineadas al nivel 3 del Modelo de Producción Estadística (MPE), se cumplieron en los tiempos establecidos, de acuerdo con el cronograma aprobado para la producción de las CSE 2021-2023.



ELABORADO POR:	REVISADO Y APROBADO POR:
<p>Miembro de Equipo Unidad de Gestión de Análisis de Síntesis</p> <p>Nombre: Nikole Pepinós</p>	<p>Responsable de la Gestión de Análisis de Síntesis</p> <p>Nombre: Henry Valdiviezo</p>



@InecEcuador



@ecuadorencifras



@ecuadorencifras



INECEcuador