

Informe de la Fase de Procesamiento

Cuentas Satélite de
Educación (CSE)
2018-2019

Septiembre, 2021

Índice

1. Introducción	3
2. Objetivo	3
3. Desarrollo	4
3.1. Criticar e integrar la base de datos	4
3.1.1. Ejecutar la crítica de la información recolectada	5
3.1.2. Ejecutar la integración/sincronización de la base de datos desde las distintas fuentes planificadas	5
3.1.3. Aplicar los controles de confidencialidad en el procesamiento	5
3.2. Clasificar y/o codificar	5
3.2.1. Clasificar y codificar la base de datos	5
3.3. Validar e imputar	5
3.3.1. Validar la base de datos	6
3.3.2. Aplicar métodos de imputación a la base de datos	8
3.4. Derivar nuevas variables y unidades	8
3.4.1. Calcular variables derivadas	8
3.5. Finalizar tabulados e indicadores de las CSE	8
3.5.1. Compilar las bases de datos derivadas del proceso	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. Anexo	10

1. Introducción

El Instituto Nacional de Estadísticas y Censos (INEC) dentro de su programación estadística plurianual y en el marco del Sistema Nacional de Planificación viene elaborando 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 Unidad de Gestión de Análisis de Síntesis (GASIN), planificó la construcción de las Cuentas Satélite de Educación (CSE) periodo 2018-2019 que serán publicadas en diciembre del presente año 2021.

Es importante destacar que las actividades de construcción 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 modelo de producción estadística que se aplican a la construcción de las Cuentas Satélite son las siguientes:

1. Planificación
2. Diseño
3. Recolección
- 4. Procesamiento**
5. Análisis
6. Difusión
7. Evaluación

La **fase de procesamiento** comprende un conjunto de actividades recursivas 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 recursiva que puede repetirse varias veces hasta obtener bases de datos intermedias depuradas y con información de calidad para la generación de resultados en la fase de Análisis¹.

De este modo, el presente documento recoge las actividades realizadas durante la fase de procesamiento de las CSE, como: las codificaciones aplicadas a la cobertura de las instituciones del sector de la educación; las agrupaciones realizadas en función de diferentes clasificadores; diferentes métodos de crítica, validación e integración aplicada a la base de datos; y, los métodos e insumos utilizados para construir las CSE.

2. Objetivo

- Documentar las actividades realizadas en la fase de procesamiento como parte de la construcción de las Cuentas Satélite de Educación 2018-2019.

¹ Modelo de Producción Estadística del Ecuador (MPE). 2016. INEC

3. Desarrollo

La CSE es una operación estadística de síntesis que utiliza una gama de información de estadística de base como son los registros administrativos, registros estadísticos u otra información previamente publicada como las encuestas, tablas de síntesis del BCE, MINEDUC, MIES, Senescyt, etc.

El mayor esfuerzo de una operación estadística de síntesis, no se concentra en la recolección de información, tal como sucede con un proceso típico de levantamiento de información a través de una encuesta o un censo, sino que su intensidad de trabajo 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 o recolectada.

Los procesos y actividades de esta fase se ejecutan acorde a las buenas prácticas estadísticas que buscan generar datos con características de precisión, confiabilidad, coherencia y comparabilidad.

Por ello, una vez culminada la fase de recolección, se inicia con el desarrollo de las actividades programadas en la fase de procesamiento en función de las siguientes actividades:

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

En esta línea, para realizar las actividades conforme al nivel 3 del MPE, en la siguiente sección se describen las actividades realizadas por el equipo técnico de la Unidad de Gestión de Análisis de Síntesis (GASIN) durante la fase de procesamiento de las CSE.

Esta fase se divide en siete procesos generalmente secuenciales, pero que también pueden ejecutarse en procesos paralelos e iterativos.

3.1. Criticar e integrar la base 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 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).

3.1.1. Ejecutar la crítica de la información recolectada

Esta actividad se ejecuta como parte del proceso del punto 3.3.2 “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

Esta actividad se ejecuta como parte del proceso del punto 3.3.2 “Aplicar métodos de imputación a la base de datos”.

3.1.3. Aplicar los controles de confidencialidad en el procesamiento

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 2 de Diseño y se encuentra en la siguiente ubicación:

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\2_Dis\2.1_Dis_prod\2.1.3_Proc_info_confid

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 de las cuentas satélite, 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 construidas 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

Esta actividad se ejecuta como parte del punto 3.3.2 “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 son comparados 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

Para llegar a este paso se debe haber concluido previamente con la crítica, integración, codificación, clasificación, validación e imputación. La validación implica analizar y revisar exhaustivamente cada base de datos utilizada como fuente de datos primaria (estadísticas de base) de las CSE.

Las bases de datos objeto de análisis comprenden cada una de las fuentes de información definidas en la fase de la recolección. Por ello, en este punto se analizan las bases de datos por separado o de forma individual según la información que se haya recolectado, es importante recalcar que todas las bases de datos que se recolectan no tienen la misma estructura 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
- Base de datos del Ministerio de Inclusión Económica y Social
- Base de datos de SECOB
- Base de datos de Patronato San José
- Base de datos de Secretaría de Educación del Distrito Metropolitano de Quito
- Base de datos del Gobierno Local
- Base de datos del Institutos Técnicos y Tecnológicos (ITT)

Las bases de datos que se analizan en el sector privado y público poseen información sensible de micro dato, por tanto además de procesar información

consistente y precisa debe estar sujeta a los cuidados respectivos de confidencialidad de la información. Cada base de datos al final del análisis de validación debe estar conformada por un conjunto de 24 variables homologadas, las cuales posteriormente confluyen en una base de datos unificada final.

Las 24 variables que conforman las bases de datos individuales son: ejercicio, tipo, codigo_N1, codigo_N6, descr_codigo_N6, id_registro, cod_entid_inst, descr_entid_inst, cod_programa, descr_programa, cod_proyecto, descr_proyecto, cod_actividad, descr_actividad, cod_provincia, descr_provincia, cod_canton, descr_canton, part, descr_part, cod_CN, devengado, fuente y observaciones.

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 Desagregación Territorial
- Base de datos de deflatores
- Bases de datos de Equilibrios Conexos
- Bases de Programas del Ministerio de Educación

Es este punto, se realiza la elaboración, procesamiento, validación y análisis de la matriz de distribución de los niveles de enseñanza del sector público.

Así mismo, en función de las mejoras metodológicas planificadas para la elaboración de las CSE 2018-2019, como son: desagregación de la enseñanza superior en tercer y cuarto nivel y desagregación de indicadores a nivel territorial, se ejecutan procesos similares de validación de la información.

Para implementar las mejoras, se analizan las diferentes fuentes de información que permitirán alcanzar los objetivos planteados en las mejoras. Por ejemplo, para la desagregación de los niveles de la enseñanza superior se analizan las bases de datos de cobertura de establecimientos de enseñanza, base de datos de alumnos y docentes del MIES, MINEDUC Senescyt y encuestas con la finalidad de obtener la desagregación para el sector público y privado, esta validación y contraste de fuentes permitirán obtener la nueva desagregación de la enseñanza superior en las CSE.

De la misma forma, se elaboran base de datos de deflatores para cada unidad institucional que conforman las CSE, es decir el nivel más desagregado. Con el objetivo de obtener información a precios constantes.

Finalmente, se construyen las bases de datos de los equilibrios conexos a precios corrientes, es decir aquellos que contribuyen a la prestación de servicios de educación, tales como: fabricación de uniformes, textos, transporte estudiantil, entre otros.

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

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\5_Proc\5.3_Valid_imput\5.3.1_Valid_bas

3.3.2. Aplicar métodos de imputación a la base de datos

En esta sección, se elaboran informes referentes a las bases de datos construidas para el sector público, privado y otros como los procesos de construcción de matrices de distribución, metodología de desagregación territorial, etc. Los informes generados son el soporte y respaldo de la implementación de procesos orientados a satisfacer las mejoras metodológicas planificadas para la construcción de las CSE.

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

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\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

Esta actividad se ejecuta como parte del punto 3.5.1 “Compilar las bases de datos derivadas del proceso”.

3.5. Finalizar tabulados e indicadores de las CSE

Una vez finalizado el análisis de las bases individuales (mencionadas en el punto 3.3.1) y a partir de las 24 variables homologadas en cada base individual, éstas 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

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:

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\5_Proc\5.7_Finali_archiv_dat\5.7.1_Compil_bas_dat\1_BUCSE2018-19

Así mismo, adicional a la base unificada, se detallan y generan todos los demás insumos necesarios para la generación de las bases intermedias como: nomenclatura actualizada de las CSE, matrices de distribución de los niveles de enseñanza público y privado, matrices de desagregación territorial, matrices de distribución de los programas sociales del sector público, matrices de distribución de la enseñanza superior, deflatores, matriz de equilibrios conexos. Los documentos verificables incluyendo la base unificada se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\5_Proc\5.7_Finali_archiv_dat\5.7.1_Compil_bas_dat\2_Bases_trabajo

Como se mencionó, a partir de la base unificada y los demás insumos necesarios, se generan bases de datos intermedias que son los insumos principales para derivar las estadísticas de síntesis y los tabulados. Los documentos verificables de este punto se encuentran en la siguiente dirección de la carpeta compartida de la Unidad:

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\5_Proc\5.7_Finali_archiv_dat\5.7.1_Compil_bas_dat\3_Resultados

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 el efecto se construyen scripts en el programa estadístico R y se usan plantillas determinadas para cada tipo de tabulado como:

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

Adicionalmente y como parte de la documentación, se elaboran fichas técnicas de los indicadores más relevantes de la operación estadística. Finalmente, se realizan 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 la Difusión.

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

R:\CGTPE\DECON\AS\CS_MPE_2021\CSE_2018-19\5_Proc\5.7_Finali_archiv_dat\5.7.2_Compil_prod_ant

4. Anexo

Actividades de la fase de procesamiento que no aplican en la elaboración de las CSE

El cronograma de actividades para construir las CSE 2018-2019, está elaborado según Modelo de Producción Estadística (MPE) a nivel de desagregación por actividad (N3).

Por tanto, hay actividades a nivel 3 de la fase 5 de Procesamiento que no se aplican (N/A) durante la construcción de las Cuentas Satélite de Educación, esto debido a la particularidad de las operaciones de síntesis respecto de las encuestas y/o registros administrativos que son el marco de referencia de la estructura de actividades y fases del MPE. De este modo, las actividades que no se ejecutan en las CSE son las siguientes:

- 5.1.2 Digitalar la información
- 5.4.1 Calcular unidades derivadas (agregar o separar unidades)
- 5.5.1 Ajustar los factores expansión
- 5.6.1 Analizar la representatividad de los principales indicadores y tabulados

Es importante reiterar que las operaciones de síntesis utilizan registros administrativos, encuestas publicadas u otro tipo de información para realizar cálculos que generalmente tienden a resumir microdatos bajo determinadas metodologías como es el caso de las estadísticas de síntesis; por otro lado, las actividades que no se ejecutan están vinculadas con operaciones de campo (encuestas) y como tal no se reportan como actividades para elaborar las CSE.

FIRMAS DE APROBACIÓN		
ELABORADO POR:	REVISADO POR:	APROBADO POR:
<p>Miembro de Equipo Unidad de Gestión de Análisis de Síntesis</p> <p>Nombre: María Dolores Robalino</p>	<p>Jefe de Unidad de Gestión de Análisis de Síntesis</p> <p>Nombre: Henry Valdiviezo</p>	<p>Director de Estadísticas Económicas</p> <p>Nombre: Darío Vélez</p>

