



Buenas cifras,
mejores vidas



Documento del proceso y metodología de las distintas etapas de validación a las cuales debe ser sometido la Encuesta Estructural Empresarial (ENESEM)

Plan de Mejoras Nro. 26

Mayo, 2021



www.ecuadorencifras.gob.ec



sembramos
Futuro

Lenin





Tabla de contenidos

Tabla de contenidos	2
Lista de gráficos	2
1. Introducción	3
2. Objetivo	3
3. Etapas de validación	3
3.1 Fase de recolección:.....	4
3.1.1 Proceso de Pre-crítica.....	4
3.2 Fase de procesamiento.....	5
3.2.1 Proceso de Crítica e integración de la base de datos.....	5
3.3 Fase de Clasificación y codificación	6
3.4 Fase de Validación y Estimación	7
3.4.1 Proceso de Detección de inconsistencias.....	7
3.4.2 Proceso de Corrección de inconsistencias	8
3.4.3 Proceso de Generación de datos finales	11
3.4.4 Proceso de Imputación de valores	11
4. Flujo grama del proceso de validación	12
5. Conclusiones.....	14
6. Firmas de responsabilidad.....	15
7. Anexo: Malla de Validación (documento adjunto)	

Lista de gráficos

Gráfico 1 Flujo grama del proceso de validación de la ENESEM	13
--	----



Documento del proceso y metodología de las distintas etapas de validación a las cuales debe ser sometido la Encuesta Estructural Empresarial (ENESEM)

1. Introducción

La recopilación de información estadística de las empresas ha evolucionado desde el año 1955, con la ejecución del Primer Censo Industrial del Ecuador. Con el transcurso del tiempo la investigación en este ámbito ha presentado mejoras sustanciales en función de las necesidades de los usuarios de la información, la innovación tecnológica, la mejora de las prácticas estadísticas del INEC y las recomendaciones internacionales en la materia. En la actualidad, a fin de atender esta necesidad de información estadística estructural, el INEC produce con periodicidad anual la Encuesta Estructural Empresarial (ENESEM), investigación que antes del 2016 se abordaba mediante las Encuestas Industriales (Manufactura, Comercio y Servicios).

La Encuesta Estructural Empresarial (ENESEM), tiene como objetivo producir información estadística, sobre la estructura económica y la producción de las empresas grandes y medianas del Ecuador la cual genera indicadores económicos de importancia como: Producción, Consumo intermedio, Valor agregado, Personal Ocupado, entre otros.

En este contexto, tomando la recomendación del Plan de Mejoras Nro.26, la cual expone la necesidad de “Definir efectivamente los procesos de solución de problemas de procesamiento al finalizar el operativo de campo y la propuesta del módulo de inteligencia de negocios ayudará en este punto”. A continuación, se describe el proceso de validación al que la información de la ENESEM es sometida.

Con este proceso se plantea la optimización de los tiempos mediante un debido proceso de validación y verificación de la información; con el fin de obtener una base de datos final consistente para la generación de resultados.

2. Objetivo

Diseñar un documento de procesos y establecer la metodología de las distintas etapas de validación a las cuales debe ser sometida la información levantada por la ENESEM.

3. Etapas de validación

Las instancias de validación y verificación a las cuales debe ser sometida la información, se encuentran detallados en la fase de recolección, procesamiento y análisis de los datos de acuerdo con las mallas de validación establecidas.

Previo al levantamiento de la información, se realiza la planificación de la validación, en donde se establece cronogramas con fechas estipuladas para que cada actividad sea



cumplida, además se establecen responsables para monitorear el avance de cada una de ellas.

Durante el levantamiento de información, la validación y verificación se presenta en el proceso de pre-crítica, crítica y codificación, en donde los encuestadores y los revisores de calidad analizan la información, y por medio de las alertas que arroja el aplicativo web informático INFOCAPT, se puede tomar acciones oportunas para corregir los posibles errores.

En la fase de procesamiento se incluye la identificación de errores, revisión de las observaciones y envío de estos a los responsables zonales para la verificación, de la información cargada, de forma que se obtenga una base de datos conforme a parámetros de calidad y confiabilidad establecidos.

Las fases de Procesamiento y Análisis se realizan paralelamente, el análisis comparativo de la información permite detectar datos atípicos y determinar el comportamiento de las variables investigadas.

Los detalles del proceso de validación se describen a continuación:

3.1 Fase de recolección:

En esta fase se realiza el levantamiento de la información por parte de los encuestadores que mediante el acercamiento con el informante se revisa y ordena la información recolectada, dentro de esta fase existe un proceso de suma importancia para la validación y verificación de la información que es el proceso de pre-crítica que se describe a continuación:

3.1.1 Proceso de Pre-crítica

El Proceso de pre-crítica consiste en realizar la validación, ordenamiento y verificación de la información levantada con criterios analíticos y lógicos que establezcan la consistencia de la información levantada, para ello, la ENESEM realiza el proceso de pre-crítica a través del sistema del INFOCAPT, dado que el aplicativo genera mensajes de alerta cuando existen posibles errores en el registro de la información. Para esto se trabaja en conjunto con DIRAD, con mallas de validación (archivo anexo a este documento) generadas por el equipo de la Unidad de Gestión de Estadísticas Estructurales (GESE), a fin de detectar errores o inconsistencias de la información.

Para garantizar un sistema de alerta eficiente se construye y actualiza el aplicativo web INFOCAPT, que está sujeto a revisión, actualización y pruebas para su correcto funcionamiento. Las pruebas permiten detectar y mitigar los errores en el diseño y funcionalidad del aplicativo en los cuatro módulos de usuario que lo integran: módulo de informante, módulo de crítica-codificación, módulo de revisor y módulo de



administrador. Por tanto, el aplicativo automatiza los procesos de recolección, crítica, revisión y procesamiento de la información levantada.

Para facilitar el manejo de la herramienta se dispone de un manual denominado “Manual del entrevistador, crítica y codificación ENESEM” que detalla las instrucciones de llenado, facilitando la comprensión del informante. El acceso al aplicativo se realiza con conexión a internet.

Para seguridad y control de la información el usuario dispone de una clave y contraseña única para el ingreso al sistema, que debe ser personalizada para garantizar la confidencialidad en la entrega de información.

3.2 Fase de procesamiento

La fase de procesamiento corresponde a la depuración de datos, generación de resultados estadísticos y su preparación para el análisis y difusión. De ser necesario los procesos de esta fase pueden repetirse varias veces, hasta obtener un producto final para la fase de análisis

En esta fase, la ENESEM ejecuta actividades de crítica, codificación, correcta utilización de clasificaciones y/o codificaciones, validaciones parciales y finales de la base de datos y en caso de ser necesario se utilizan métodos de imputación para garantizar precisión y coherencia de la data recopilada en campo.

3.2.1 Proceso de Crítica e integración de la base de datos

Posterior a esto, se inicia el proceso de crítica-codificación, donde se revisa la calidad de la información levantada, se detallan las observaciones registradas y se hace una **validación** de la consistencia de la información, de manera inmediata se realiza el proceso de revisión, ejecutado por los revisores de calidad, a fin de revisar los flujos de la encuesta, consistencia, cruce de variables y duplicidad de información; apoyados en los manuales, se busca disminuir los errores y dar soluciones oportunas que garanticen la calidad de la información.

En caso de detectar inconsistencias en los datos, el personal de crítica toma contacto telefónico con los informantes, según sea el caso (información económica, ambiental, TIC, etc.), o a su vez consulta al encuestador, quien podría solventar la observación; caso contrario, se toma contacto nuevamente al informante para solicitar notas aclaratorias a fin de verificar la información proporcionada.

Una vez culminado el proceso de crítica y revisión, la Dirección de Registros Administrativos (DIRAD) integra bases de datos primarias en una sola, para luego transferirla a la Unidad de Gestión de Estadísticas Estructurales (GESE).

A continuación, se describen las actividades realizadas para la generación y revisión de las bases de datos integrada:



1. La Dirección de Estadísticas Económicas, establece con la Dirección de Registros Administrativos, un cronograma de entrega de bases de datos parciales a fin de realizar validaciones preliminares de información.
2. Una vez finalizado el proceso de levantamiento de información en campo, digitación, crítica y codificación, DECON solicita a DIRAD la base de datos final.
3. Se verifica que consten todas las variables y casos de estudio, en caso de encontrar inconsistencias se reporta a DIRAD las novedades y se solicita nuevamente la base de datos para posterior proceso de validación.
4. A la base final remitida por DIRAD la Unidad de Gestión de Estadísticas Estructurales (GESE) se encarga de realizar el tratamiento de las variables incluidas las variables del componente 1 demanda laboral, que consiste en su validación, etiquetado, depuración y análisis para su posterior publicación de los resultados obtenidos.

3.3 Fase de Clasificación y codificación

En este proceso se clasifican y codifican las bases de datos integradas utilizando las clasificaciones estadísticas con estándares internacionales e instrumentos lógicos, adaptados a la realidad nacional, que agrupan, ordenan y categorizan fenómenos u objetos de la realidad económica y social del país.

La codificación consiste en transformar la información recopilada, con el instrumento respectivo a códigos numéricos apropiados, coherentes y previamente establecidos; es así como, la ENESEM utiliza nomenclaturas internacionales estandarizadas que permiten clasificar las actividades, productos, servicios y ocupaciones que ofrecen las empresas investigadas por la encuesta. A continuación, se detalla las clasificaciones utilizadas:

- **Clasificación Industrial Internacional Uniforme (CIIU) de actividades económicas Rev. 4.0.**

Es la clasificación internacional de referencia de las actividades productivas. Su propósito principal es ofrecer un conjunto de categorías de actividades que se pueda utilizar para la reunión y difusión de datos estadísticos de acuerdo con esas actividades (ONU, 2009).

En este marco, la clasificación CIIU permite clasificar uniformemente las actividades o unidades económicas de producción (empresas o establecimientos) de la ENESEM.

- **Clasificación Nacional de Ocupaciones (CIUO) Rev. 08.**

Es un sistema organizado de información sobre las diferentes ocupaciones de la población económicamente activa y se aplica para codificar la información del personal ocupado (capítulo 5) de la ENESEM.



Facilita un marco para el análisis, la agregación y la descripción de los contenidos del trabajo en una serie de grupos definidos claramente en función de las tareas que intervienen en cada empleo, la unidad de clasificación de la CIUO son las ocupaciones no títulos profesionales. (OIT, 2005)

- **División Política Administrativa (DPA) versión 2019**

La ENESEM utiliza como insumo el DPA 2019, con el fin de homologar la clasificación de las provincias del país, además este documento es una guía para la clasificación de los cantones y parroquias (urbanas y rurales) y sus respectivas cabeceras o sedes administrativas, según la desagregación de la operación estadística lo demande.

3.4 Fase de Validación y Estimación

La validación y estimación consiste en examinar los datos levantados dentro de la encuesta para identificar y determinar los posibles problemas, inconsistencias y errores, como los valores atípicos, la falta de respuesta, errores en la codificación, mala digitación en los valores, etc.

La validación de la información de la encuesta se basa en 3 procesos que consisten en:

1. Proceso de Detección de inconsistencias.
2. Proceso de Corrección inconsistencias.
3. Proceso de Generación de datos finales.
4. Proceso de Imputación de valores.

3.4.1 Proceso de Detección de inconsistencias

Las principales actividades que se realizan en el proceso de detección de inconsistencias se describen a continuación.

Proceso de Planificación

El proceso de planificación se lo realiza antes del levantamiento de la información y corresponde a la elaboración de un cronograma de ejecución de actividades y responsables. Esto permite llevar un estricto control de la ejecución de las actividades y los tiempos estipulados de ejecución, garantizando que las actividades se realicen de la mejor manera. Dentro de las actividades están el diseño de las mallas de validación, diseño de la sintaxis validación, la coordinación con DIRAD para las entregas de las bases de datos preliminares y finales, entre otros.



Diseño de las mallas de validación

Una malla de validación permite revisar de manera automática la consistencia de la información proveniente de cada encuesta y los rangos de las respuestas, identificando de manera automática cuando un valor o respuesta específica no corresponda a un rango, por lo tanto el diseño de las mallas de validación constituye un paso fundamental en el proceso de detección de inconsistencias, pues emitirá alertas cuando se detecten errores en la carga de información y a su vez permite una pre-crítica de la información registrada, dentro de las mallas de validación está el control de sumas, validación entre variables etc.

Comparación de la información

En el levantamiento, es importante comparar los casos de la base de datos con la información del reporte de cobertura realizado por las Coordinaciones Zonales (directorio de empresas de la encuesta), pues permite verificar algún caso de inconsistencia de información.

En definitiva, para la identificación de errores, se debe realizar un estricto análisis de la base de datos en cada validación, para lo cual se debe utilizar el archivo de sintaxis de validación desarrollado en el programa estadístico SPSS. Una vez generados los resultados con la sintaxis de validación, se identifica los casos que presentan inconsistencias, como: información faltante, datos atípicos, encuestas con posibles errores de digitación, errores en los controles de suma, errores en las igualdades, errores en los cruces de variables etc.

Es importante mencionar que se debe realizar una comparación entre la BDD y el Directorio de Empresas para validar importante información tal como: Ubicación de las empresas, número de trabajadores, tamaño de la empresa, el valor de las ventas y la actividad económica a la que pertenece la empresa. Dentro de la detección de errores debe estar el tema de la zonificación, para lo cual se debe utilizar la codificación actual. En esta validación se debe verificar la correcta zonificación de la provincia, cantón, parroquia, zona, sector y manzana. De todo este proceso se debe identificar todos los errores y se deben enviar a los distintos responsables zonales para su posterior corrección.

3.4.2 Proceso de Corrección de inconsistencias

A continuación, se describen las principales actividades realizadas en este proceso de validación

Revisión de observaciones



Se realiza una revisión de las observaciones registradas en el proceso de crítica de los errores o inconsistencias presentadas en el aplicativo, para determinar si dichas observaciones cuentan con el rigor estadístico, además de obedecer a la realidad de cada empresa y proveer de una explicación lógica a la alerta detectada por parte del aplicativo.

Sintaxis de Validación

La sintaxis de validación es un proceso que asegura la entrega de datos limpios y de calidad, además comprueba la integridad y validez de los datos. La sintaxis de validación garantiza que los datos cumplan requisitos y parámetros de calidad.

Actualizar la sintaxis de validación es un proceso fundamental pues esto permitirá ejecutar la rutina en cada corte de base de datos entregada por DIRAD. La sintaxis permitirá realizar un control de sumas, validación del flujo de la encuesta y consistencia de los datos mediante la aplicación de la malla de validación. Se pone como archivo anexo a este documento la malla de validación.

Para la revisión de la base se desarrolla una sintaxis de validación apoyados en la herramienta estadística SPSS en la cual se desarrollan los siguientes ejemplos de validación:

- **Validación de sumas:** Para la validación de sumas se desarrolla la siguiente sintaxis de ejemplo donde el programa SPSS arrojará un error en el caso de que el valor de la suma no sea el correcto.

Ejemplo de sintaxis de validación de sumas:

***v1001 Total de Ingresos.

IF (v1001 = sum.1(v1002, v1012, v1029, v1039)) v1001c=1.

IF (v1001 ~= sum.1(v1002, v1012, v1029, v1039)) v1001c=2.

VALUE LABELS v1001c 1 'Bien' 2 'Error'.

EXECUTE.

VARIABLE LABELS v1001c 'Control de suma Total de Ingresos'.

FREQUENCIES v1001c.

- **Validación de cruce de variables:** Para la validación de cruce de variables se desarrolla la siguiente sintaxis de ejemplo donde el programa SPSS arrojará un error en el caso de que la suma de valores registrados en el capítulo A no cuadre con la suma de valores del capítulo B.



Ejemplo de sintaxis de validación de cruce de variables:

* La suma de valores registrados en las filas (112 + 113 + 114) debe ser igual a la suma de valores registrados en el Capítulo 1, filas (4 + 6).

```
COMPUTE cr_112_113_114 = (SUM.1(v2001, v2002, v2003) = SUM.1(v1004, v1006)).  
EXECUTE.  
RECODE cr_112_113_114 (0=2).  
EXECUTE.  
VARIABLE LABELS cr_112_113_114 'La suma de valores registrados en las filas (112 +  
113 + 114)'+  
' debe ser igual a la suma de valores registrados en el Capítulo 1, filas (4 + 6)'.  
VALUE LABELS cr_112_113_114 1 'Bien' 2 'Error'.  
FORMATS cr_112_113_114 (F1.0).  
EXECUTE.  
FREQUENCIES cr_112_113_114.
```

- **Validación de igualdades:** Para la validación de igualdades se desarrolla la siguiente sintaxis de ejemplo donde el programa SPSS arrojará un error en el caso de que el valor registrado en el capítulo A no sea igual al valor registrado en el capítulo B.

Ejemplo de sintaxis de validación de igualdades:

* El valor registrado en la fila 117 debe ser igual al valor registrado en el Capítulo 1, fila 2, código 1002 de Ingresos de actividades ordinarias.

```
COMPUTE cr_117 = (v2006 = v1002).  
EXECUTE.  
RECODE cr_117 (0=2).  
EXECUTE.  
VARIABLE LABELS cr_117 'El valor registrado en la fila 117 debe ser igual al valor'+  
' registrado en el Capítulo 1, fila 2, código 1002 de Ingresos de actividades ordinarias'.  
VALUE LABELS cr_117 1 'Bien' 2 'Error'.  
FORMATS cr_117 (F1.0).  
EXECUTE.  
FREQUENCIES cr_117.
```

Remitir Observaciones

Una vez que se ha realizado la validación con la ayuda del programa estadístico SPSS y la identificación de los distintos errores encontrados en la Base de Datos se debe remitir las observaciones detectadas a través de una rutina de validación a las Coordinaciones Zonales para su corrección o justificación, tanto en las validaciones parciales, como en



la validación final; los ajustes se hacen en el aplicativo web. Se debe realizar las validaciones por cada entrega de la base de datos de DIRAD, en las fechas estipuladas.

Corregir los errores

Se debe corregir los errores encontrados en la validación de la base de datos mediante el ajuste de los valores en el aplicativo web, contrastando su coherencia con valores relacionados entre los distintos capítulos.

3.4.3 Proceso de Generación de datos finales

En este apartado se describe las acciones que se ejecutan a fin de generar la base de datos final.

Base de datos Final

Una vez realizadas todas las correcciones o justificaciones de la información con alerta, se solicita a DIRAD la base de datos final a fin de verificar las correcciones realizadas y posteriormente generar los resultados de la encuesta.

Desarrollo de etiquetado

Dentro de esta actividad se desarrolla la sintaxis de etiquetado, preparación ajuste y diseño de las variables que se procederán a publicar en la base de datos final.

3.4.4 Proceso de Imputación de valores

De acuerdo con las recomendaciones internacionales “La falta de respuesta del elemento o la falta de respuesta parcial ocurre cuando la unidad tomada como muestra no respondió a todas las preguntas pertinentes, sino a parte de ellas. Pueden darse casos en los que un encuestado haya respondido a todas las preguntas pero que algunas de las respuestas sean ilógicas, o bien que existan incoherencias entre algunas de las preguntas contestadas por el encuestado”, por lo tanto “La presencia de este tipo de falta de respuesta del elemento o de datos inválidos en el conjunto de datos afecta en última instancia a la calidad de los resultados de la encuesta” (ONU, 2008).

La presencia de datos faltantes es la situación a la que permanentemente se enfrentan investigadores y tomadores de decisiones, lo ideal es disponer de un archivo de datos completos (CEPAL, 2007). Por lo cual, es importante precisar que la ENESEM realiza un proceso de imputación cuando se presentan datos faltantes completos o parciales, este proceso se basa en los registros administrativos empresariales e información histórica de la unidad de análisis para no afectar la calidad y representatividad de los resultados de la encuesta.



Con lo antes expuesto, la ENESEM realiza un proceso de imputación para datos faltantes completos o parciales, basado en los registros administrativos empresariales e información histórica de la unidad de análisis, para mejorar la calidad y representatividad de los resultados de la encuesta. Se busca así mitigar el impacto del incremento de la tasa de sub-cobertura de campo, ocasionado por el crecimiento de novedades no efectivas (rechazos, no ubicadas, entre otras.)

Con base en las recomendaciones y metodologías internacionales, para elegir un caso de imputación (empresa) se consideró los siguientes criterios:

1. La representatividad (ventas) de las empresas dentro de su dominio de estudio, que rechazaron la encuesta o no fueron ubicadas
2. Disponibilidad de los valores totales de ventas y costos en los registros administrativos del SRI, y del personal ocupado en el registro del IESS.
3. Empresas con información que fue declarada en rondas anteriores de la ENESEM, de preferencia en el año más reciente (2018); esto para observar la estructura de las sub-cuentas y poder extrapolar al registro faltante.

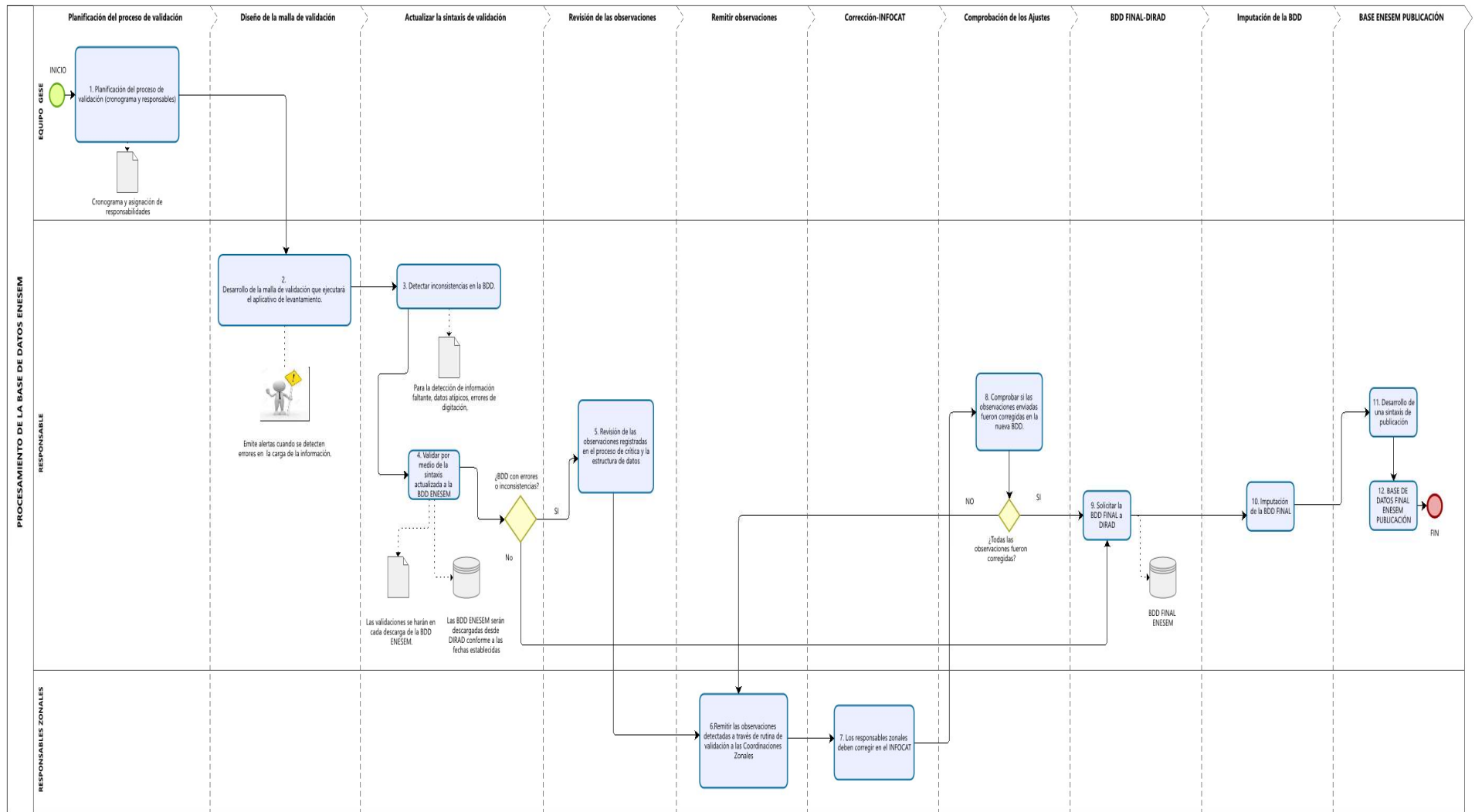
El proceso de imputación está basado en el uso de “Datos duros” a través del aprovechamiento de registros administrativos (INEGI, 2021) utilizando la información reportada por el Directorio de Empresas y Establecimientos Económicos 2019 (DIEE2019), que insume registros administrativos del SRI e IESS.

El procedimiento consiste en completar las observaciones con datos faltantes, considerando los totales de las variables ventas y costos del registro administrativo del SRI, y del personal ocupado del registro del IESS. Para la imputación de los valores desglosados de las variables, se replica la estructura de los valores reportados en años anteriores, preferentemente del año más reciente.

4. Flujo grama del proceso de validación

Para resumir el proceso de validación de la encuesta ENESEM, se desarrolla un flujo grama representado en el gráfico 1, el cual empieza con la planificación de la validación y termina en la BDD final.

Gráfico 1 Flujo grama del proceso de validación de la ENESEM



Fuente: Validación ENESEM 2019

Elaboración: Gestión de Estadísticas Estructurales

5. Conclusiones

- En este documento se detallan las etapas de validación a los que es sometida la información levantada en la ENESEM.
- Los procesos de validación se realizan antes, durante y después del levantamiento de información por lo que se determina que antes del levantamiento se debe planificar un cronograma de actividades, que incluye el desarrollo de la malla de validación, desarrollo y prueba del correcto funcionamiento del aplicativo web INFOCAT, y desarrollo de la sintaxis de validación de la base de datos.
- En el levantamiento se debe realizar una validación por cada corte de la base de datos, en este proceso se identifican los errores o inconsistencia de la información por medio de una sintaxis previamente desarrollada, posterior a esto se realiza una revisión de las observaciones registradas en el proceso de crítica de las inconsistencias presentadas en el aplicativo, para determinar si dichas observaciones cuentan con el rigor estadístico, este proceso se repite las veces que sean necesarias hasta conseguir un producto final que es la base de datos para publicación.
- Después del levantamiento se hace una validación final de la BDD, en esta validación no deben existir errores o deben ser mínimos, dado que todos los errores se fueron solventando a medida que se realizaban las validaciones preliminares.



6. Firmas de responsabilidad

ELABORADO POR:	Daniel Rodríguez Miembro del equipo	
	Jonathan Quijia Miembro del equipo	
REVISADO POR:	Roberto Chaves Jefe de área	
	Darío Vélez Director de Estadísticas Económicas	
APROBADO POR	David Sánchez Coordinador General Técnico de Producción Estadística	



 **INEC** | Buenas cifras,
mejores vidas



@ecuadorencifras



@ecuadorencifras



@InecEcuador



t.me/euadorencifras



INEC/Ecuador



INECEcuador



INEC Ecuador