

Informe de validación de la base de datos - Encuesta Nacional de Edificaciones (ENED) 2020

Agosto, 2021

Tabla de contenido

1. Introducción.....	3
2. Objetivo	3
3. Desarrollo	3
3.1. Descarga de la base de datos y análisis preliminar	3
3.2. Identificación de errores.....	4
3.3. Corrección de errores.....	5
3.4. Comprobación de ajustes realizados.....	6
3.5. Validación final y corrección	6
3.6. Proceso de imputación.	8
4. Generación de la base - Formato publicación.....	8
5. Conclusión.....	9

Índice de Tablas

Tabla 1. Número de casos erróneos encontrados en la base de datos ENED 2020, por tipo de validación	4
Tabla 2 Número de casos, que dado su naturaleza no pueden ser corregido por sintaxis	5
Tabla 3 Cálculo de variables de publicación.....	8

Informe de validación de la base de datos - Encuesta Nacional de Edificaciones (ENED) 2020

1. Introducción

La Encuesta de Edificaciones (ENED) es una operación estadística de periodicidad anual que detalla los potenciales proyectos de inversión en edificaciones, a través de la información derivada de los permisos de construcción emitidos por los Gobiernos Autónomos Descentralizados Municipales (GAD) del país. Los resultados de esta encuesta constituyen una señal importante para identificar las expectativas de construcción y la demanda de inversión de los hogares.

2. Objetivo

Detallar los procedimientos considerados para la validación de la base de datos de la Encuesta de Edificaciones período 2020.

3. Desarrollo

La validación de la base de datos de la Encuesta Nacional de Edificaciones (ENED) responde a uno de los procesos esenciales dentro de las actividades de esta operación estadística definidas en la fase de procesamiento según el MPE. Es así como, para el cumplimiento del proceso de validación de la base ENED 2020 se llevaron a cabo una serie de actividades, las mismas que se detallan a continuación:

3.1. Descarga de la base de datos y análisis preliminar

Con fecha 20 de mayo del 2021, se procedió con la descarga de la base preliminar desde el Repositorio Institucional- RIBD, para posteriormente realizar una revisión integral de la base. En esta base se esperaba que se cumplan los criterios tales como: constancia de todas las variables del formulario, cumplimiento de las anchuras y tipos de variables solicitadas en la malla de validación, sin embargo, existieron algunas variables faltantes: sec1_dist_espacio_verde y sec8_aut_cedula. Se envía las observaciones a la Dirección de Registros administrativos (DIRAD).

El 28 de mayo se procede a descargar la base desde el Repositorio Institucional -RIB, nuevamente se realiza una revisión integral de la base. En esta base constan todas las variables del formulario, se cumple las anchuras y tipos de variables solicitadas en la malla de validación. Un análisis más profundo permitió detectar una gran cantidad de datos faltantes en capítulos completos.

Se resumen las observaciones en un documento denominado "Análisis_exploratorio_BDD_ENED_28_05_2021", en el que se detalla todos los datos faltantes, para que DIRAD proceda a solucionar.

El 3 de junio, se procede a la descarga de la base del Repositorio Institucional -RIB, para revisar los casos enviados a solucionar, en esta base se evidencia que se corrige una gran cantidad de los casos enviados, sin embargo, aún existe un número importante de casos.

El 11 de junio, nuevamente se procede a la descarga de la base de datos del Repositorio Institucional-RIB para revisar los casos de datos faltantes, en esta base

aún se evidencia datos perdidos, por lo cual se envía un documento denominado "Análisis_exploratorio_11-06-2021" con los casos. Finalmente, el 1 de julio del 2021 se procede a descargar la base de datos final, en el que se hace una revisión minuciosa e integral de la base y se verifica que todas las inconsistencias fueron solucionadas.

3.2. Identificación de errores

A continuación, se realizó un estricto sondeo de la base de datos, para lo cual se utilizó el archivo de la sintaxis de validación de la ENED 2020 denominado "Validacion_ENED2020_V1", cabe recalcar que la Encuesta Nacional de Edificaciones ENED 2020, es una encuesta que tiene cambios importantes en su estructura en relación con la ENED 2019. En la ENED 2020, se incluyen nuevas variables para fortalecer la encuesta y conocer con más profundidad la dinámica del sector de la construcción. Para el desarrollo de la malla de validación ENED 2020, se tomó como base la malla de validación de la ENED 2019, sin embargo, se realizó cambios importantes debido a que hay un cambio de estructura en el ENED 2020.

Una vez generados los resultados con la sintaxis de validación, se identificó los casos que presentaban inconsistencias, como: información faltante, datos atípicos, registros con posibles errores de digitación, boletas mal criticadas y en general, casos que no cumplían con la malla de validación. Se identificó observaciones que pueden ser corregidos mediante sintaxis, y observaciones que dado su naturaleza no se ha podido realizar la corrección por sintaxis y se ha procedido a solicitar las boletas.

Para resumir las observaciones, se ha elaborado una matriz de observaciones denominada "Matriz_de_observaciones_BDD-ENED2020", en la cual se resumieron todos los casos erróneos mencionados con anterioridad a fin de remitir esta información a las Coordinaciones Zonales para que los críticos puedan comparar estos datos con la información de las boletas y realizar las correcciones pertinentes.

Con fecha 06 de junio de 2021, se remitieron 911 casos distribuidos de la siguiente manera: CZ Norte (50), CZ Centro (180), CZ Litoral (350) y CZ Sur (351). En la Tabla 1 se resumen los errores encontrados y que fueron resumidos en la matriz de observaciones.

Tabla 1. Número de casos erróneos encontrados en la base de datos ENED 2020, por tipo de validación

Nº de casos	Tipo de validación
29	Debe estar seleccionado en la variable "TIPO DE OBRA" alguna codificación, y debe registrar la información en todas las secciones que correspondan.
15	La superficie del terreno no debe superar los 15.000 m2 en edificaciones de uso no residencial.
8	El área en planta baja no debe ser mayor a la superficie total del terreno

N° de casos	Tipo de validación
141	Si el tipo de obra es nueva construcción, la sección IV debe tener lleno todos los campos.
8	Si la estructura es de madera, no se puede tener cubierta de hormigón armado.
6	Si se selecciona el uso residencial edificaciones para una vivienda el número de viviendas debe ser siempre 1, para edificaciones de dos viviendas el número de viviendas debe ser 2 y para edificaciones de tres o más viviendas el número de viviendas debe ser igual o mayor a tres.
663	El número de cuartos por vivienda debe ser máximo 12.
41	El área de la sección V debe ser igual al área de la sección III.

Fuente: Encuesta Nacional de Edificaciones 2020.

Cabe mencionar que la base de datos ENED 2020, no contiene métodos estadísticos de imputación, en caso de detectar datos faltantes fueron corregidos con la información real reportada por el informante, mediante la revisión de boletas físicas.

Con fecha 11 de agosto, se elabora un documento denominado "Matriz_boletas_ENED2020", en el que se resume los casos que dado su naturaleza no pudieron ser corregidos por sintaxis. Por tal motivo, se procedió a pedir las boletas escaneadas, o en su defecto la matriz de carga. Para la coordinación litoral se procedió a solicitar 40 boletas, y para la coordinación sur, se procedió a solicitar 16 boletas. Los casos por las cuales se solicitaron las boletas se resumen en la siguiente tabla.

Tabla 2 Número de casos, que dado su naturaleza no pueden ser corregido por sintaxis

N° de casos	Tipo de validación
2	La superficie máxima del terreno para uso RESIDENCIAL es 5000 metros cuadrados
2	La superficie del terreno no debe superar los 15.000 m2 en edificaciones de uso no residencial.
36	El monto de la sección V (uso, características y financiamiento) no es igual al monto de la sección VI (Valor y origen del financiamiento)
13	El monto por área total de construcción es muy bajo.
3	Tienen totalizado la sección V (uso residencial), pero no tiene la información de los usos.

Fuente: Encuesta Nacional de Edificaciones 2020.

3.3. Corrección de errores

El proceso de revisión de las boletas por parte de las Coordinaciones Zonales consistió en lo siguiente: en caso de que la información se haya ingresado erróneamente, se corrigió directamente en el aplicativo web; mientras que, si los datos estaban digitados correctamente se detalló la respectiva justificación dentro de la matriz, y en otros casos, se solicitó las boletas escaneadas para la verificación y análisis por parte de los técnicos de Planta Central.

Las observaciones remitidas por las zonales y el detalle de los ajustes realizados en estos casos particulares se resumen en el documento denominado "Matriz_observaciones_BDD_ENE2020_respuestas".

Con respecto a las boletas que se solicitaron, se procedió a revisar minuciosamente la información registrada, y darle el tratamiento especial a cada caso. En algunos casos existieron errores de digitación, en otros casos existieron errores del aplicativo web, y en otros fue un error de la crítica y digitación. Este último caso sucedió con la coordinación litoral, quienes digitaron 30 boletas de manera errónea los montos de la sección V y la sección VI. Para solucionar esto, se procedió a solicitar las matrices de carga originales que vienen directamente desde el municipio, con este insumo se pudo solucionar estos casos. El tratamiento de cada caso se resume en el documento denominado "Matriz_boletas_ENED2020_tratamiento".

3.4. Comprobación de ajustes realizados

Con fecha 1 de julio del 2021 se procede a realizar la descarga de la última base de datos de la Encuesta Nacional de Edificaciones – ENED 2020, la cual fue descargada desde el sistema RIBD.

Consecuentemente, se procedió a verificar que las correcciones solicitadas a los técnicos responsables zonales se encuentren plasmadas en la base de datos. Tras esta verificación se identificó que todas las observaciones fueron solventadas, sin embargo, es importante mencionar que esta validación se trabajó con una base de datos preliminar, pues la base de datos final recién se cargó el 1 de julio de 2021. En esta descarga se corre la sintaxis de validación completa y se identifican más inconsistencias. Todas estas inconsistencias se detallan en un documento denominado "Matriz_observaciones_BDD_ENE2020_01-07-2021".

3.5. Validación final y corrección

Para la corrección de todas las inconsistencias, se procede a elaborar un documento denominado "sintaxis_corrección_ENED2020" desarrollado en el paquete SPSS. La gran mayoría de las observaciones se corrigen por sintaxis, excepto casos especiales que son necesarios las boletas escaneadas, para su posterior análisis por parte de los técnicos de planta central.

Tras el análisis de las boletas escaneadas, se incluyó en la sintaxis de validación el ajuste de todos aquellos casos particulares que debían ser corregidos individualmente. Cabe indicar que dentro del análisis se tuvo como prioridad respetar la información original de la boleta, la cual es emitida por el informante.

De esta manera, se actualizó la sintaxis de validación según los requerimientos de la base de datos ENED 2020, dando como resultado el producto denominado "Edificaciones2020-Validada_final_V1". Es importante señalar que, las nuevas validaciones incorporadas en el año 2019 también se integran al tratamiento del año 2020, validaciones tales como:

Sección III:

- ✓ Si el tipo de obra es nueva construcción, el área total a construirse debe ser mayor o igual a 35 m².
- ✓ Si el tipo de obra es ampliación horizontal, el área total a construirse debe ser menor o igual a 40 m², y el área para espacios verdes debe ser cero.
- ✓ Se omite la validación que restringe el área verde a 5.000 m² en edificaciones residenciales y 15.000 m² en edificaciones no residenciales, y se toma en cuenta estas restricciones únicamente para la variable de superficie.

Sección V:

- ✓ Para casos de edificaciones residenciales, el número de viviendas totales debe ser mayor o igual al número de edificaciones a construir.
- ✓ Si el tipo de obra es reconstrucción, el número de pisos debe ser igual a cero, ya sea una edificación residencial o no residencial.

Sección VI

- ✓ En la sección VI, todos los registros deben contar con al menos un tipo de financiamiento, sea estos por recursos propios o recursos provenientes de préstamos.

Sección VII:

- ✓ Si el tipo de edificación es residencial o mixta, debe tener seleccionada alguna de las opciones "Sí" o "No" en segmento de crédito, a excepción de los casos de reconstrucción.
- ✓ Si el tipo de edificación es exclusivamente no residencial o es una reconstrucción, no debe tener respuesta en segmento de crédito.
- ✓ El número de viviendas reportado en la sección V debe ser mayor o igual al número de viviendas de la sección VII.
- ✓ Si la edificación se encuentra financiada únicamente con recursos provenientes de préstamos del Banco del Estado (con bono de vivienda MIDUVI), debe registrarse como vivienda de interés social.
- ✓ Si como propietario o proyectista consta el Ministerio de Desarrollo Urbano y Vivienda (MIDUVI) o la Empresa Pública Casa Para Todos (EPCPT), debe registrar número de viviendas de interés social o público, según corresponda.
- ✓ En caso de haber seleccionado la opción "Sí" en segmento de crédito, la clasificación de los programas de vivienda debe seguir el siguiente criterio:

Programa	Financiamiento por vivienda (USD)
----------	-----------------------------------

Vivienda de interés social	hasta 71.064
Vivienda de interés público	71.065 – 91.368

Las viviendas del programa Casa son reclasificadas dentro de los programas de viviendas de interés social o público.

- ✓ Si el financiamiento por vivienda supera los 90.000 dólares, debe estar seleccionada la opción "No" en segmento de crédito.

Finalmente, se genera nuevos resultados con los ajustes realizados en la sintaxis para la ENED 2020, a fin de dar cumplimiento en su totalidad a lo estipulado en la malla de validación.

3.6. Proceso de imputación.

Dada la naturaleza de la operación, se señala que, la ENED no aplica procesos o métodos de imputación a la información recolectada y validada, considerando que el levantamiento de información se realiza respecto de la efectividad del número de permisos o licencias emitidas en a nivel de los 221 GAD municipal del país.

4. Generación de la base - Formato publicación

Una vez obtenidas las variables de validación, se procedió con la generación de la base de datos final en formato de publicación, para lo cual se desarrolló un documento denominado "Calculo_de_variables_ENED2020".

En esta sintaxis se da la forma a las variables tal como etiqueta, valores, tipo de dato, anchura, y el nivel de la variable. Se crea una serie de variables de publicación que vienen directamente de la base de datos validada y corregida, en la siguiente tabla se detalla las variables más importantes de publicación, adicional se integra las nuevas variables generadas en la investigación del periodo 2020.

Tabla 3 Cálculo de variables de publicación

Código de la variable	Nombre de la variable	Método de cálculo
CPERM	Número de edificaciones	\sum Número de edificaciones a construir
CNUVICAL	Número de viviendas por edificación calculadas	\sum Número de viviendas
NUCUCAL	Número de cuartos por vivienda	\sum Número de cuartos
NUPICAL	Número de pisos calculado	\sum Número de pisos residencial y no residencial
NUDOCAL	Número de dormitorios por vivienda	\sum Número de dormitorios
CVAE	Valor total de la edificación calculado	\sum Valor de financiamiento para edificación
CARES	Superficie a construir residencial	\sum Área residencial a construir en m2 + \sum Área de garage exclusivo para la vivienda

CARNRES	Superficie a construir no residencial	\sum Área no residencial a construir en m2 (excepto garage y parqueadero)
CARDPAR	Superficie a construir garage	\sum Área de parqueadero en m2
CAESV	Área para espacios verdes en m2	Área para espacios verdes
CSUTE	Superficie del terreno en m2	Superficie total del terreno
CARCO	Superficie a construir total	CARES + CARNRES + CARDPAR
ACINTERN	Tipo de acabados internos	Variables categóricas
INTERN	Cobertura a internet	Variables categóricas
COBERCEL	Cobertura a telefonía celular	Variables categóricas
ACCES	Acceso a la edificación	Variables categóricas
OCUPA	Ocupación de la edificación	Variables categóricas

Fuente: Encuesta Nacional de Edificaciones 2020

5. Conclusión

El proceso de validación de bases de datos es esencial dentro de toda operación estadística, ya que garantiza la calidad de los datos. Para la ENED 2020, esta actividad demandó un tiempo considerable, el cual permitió generar una base de datos consistente en información a pesar de generar variaciones fuertes en el comportamiento a nivel de todas las variables respecto del periodo 2019, situación que se presenta a causa de factores externos que afectó al sector de la construcción. La base de datos validada permitirá continuar con el proceso de análisis y elaboración de productos de publicación.

27 de agosto de 2021.

Elaborado por:	Revisado por:	Aprobado por:
Jonathan Quijia (Miembro de equipo)	Roberto Chaves (Jefa de unidad)	Darío Vélez (Director de Estadísticas Económicas)

