

Módulo de Información Económica Ambiental en Empresas ENESEM 2020.

Guía sobre el uso de la base de datos.

Septiembre, 2022

Índice de Contenidos

Introducción	5
Requisitos técnicos	5
Resumen metodológico y alcance de los datos	7
Variables de diseño de la muestra	10
Variables de la BDD del Módulo Ambiental de la ENESEM 2020 y su uso	12
Recomendaciones generales previas al procesamiento de la BDD del Módulo Ambiental de la ENESEM 2020.....	18
ANEXO: Código R Markdown 4.1.0 para la generación de la BDD de desechos peligrosos	20

Índice de Figuras

Figura 1. Ficha técnica del MA de la ENESEM 2020	7
---	---

Índice de Tablas

Tabla 1. Lista de variables identificadoras de la BDD del MA de la ENESEM 2020.	12
Tabla 2. Lista de variables de diseño de la BDD del MA de la ENESEM 2020.....	12
Tabla 3. Lista parcial de variables temáticas de la BDD del MA de la ENESEM 2020.	13
Tabla 4. Lista de variables derivadas de la BDD del MA de la ENESEM 2020.....	14

01.

Introducción y requisitos técnicos

Introducción

El Módulo de Información Económica Ambiental (MA) en Empresas (ENESEM), edición 2020, es una parte de la Encuesta Estructural Empresarial especializada en temas empresariales relacionados con el ambiente, como son: (a) gestión ambiental, (b) oferta y utilización de bienes y servicios ambientales, (c) energía, combustibles y lubricantes, (d) agua, manejo de aguas residuales y de otros residuos.

El procesamiento de los tabulados e indicadores de resultados del MA de la ENESEM 2020 se lo realizó a partir de la base de datos (BDD) generada a partir del aplicativo institucional de carga de datos para encuestas y censos (INFOCAPT) con las empresas que respondieron, generalmente, a todos los capítulos y preguntas del formulario de captura de información correspondiente.

En lo sucesivo, indicaremos los requisitos técnicos para procesar la BDD del MA de la ENESEM 2020. También haremos un resumen metodológico del MA de la ENESEM 2020, así como se indicará los formatos de entrega de la BDD, su sintaxis de procesamiento de los tabulados e indicadores publicados, el nombre y formato de sus variables, los tipos de variables y se concluirá con un acápite sobre recomendaciones generales previas a la realización de procesamiento de usuario del MA de la ENESEM 2020.

Requisitos técnicos

Para poder procesar la BDD del Módulo Ambiental de la ENESEM 2020, el usuario debe satisfacer los siguientes requisitos, tanto de conocimiento como de infraestructura, los cuales se asumirá, en lo que sigue de este documento, que se han satisfecho, al

menos en un grado aceptable o mejor. Estos requisitos son:

- (a) *Requisitos de conocimiento:* El usuario requiere tener una formación, académica y/o práctica en Estadística Descriptiva e Inferencial. También se requiere del dominio de Microsoft Excel® 2013 o posterior, sobre todo en manejo de fórmulas aritméticas básicas, filtros, ordenamientos y gráficos. Además, es altamente deseable el conocimiento de plataformas o entornos de procesamiento estadístico, como SPSS v17 o posterior, R v4.0.3 o posterior, y Python 3.7 o posterior.
- (b) *Requisitos de infraestructura:* El usuario requiere tener instalados, al menos, Microsoft Excel® 2013 o posterior, así como R v4.0.3 o posterior, de preferencia junto al entorno gráfico auxiliar de procesamiento para R denominado *RStudio* v1.3 o posterior, en el cual se desarrolló la mayor parte de la sintaxis de tabulados e indicadores de publicación de la ENESEM 2020.

02.

Resumen metodológico y alcance de los datos

Resumen metodológico y alcance de los datos

Figura 1. Ficha técnica del MA de la ENESEM 2020

Ficha técnica del Módulo Ambiental ENESEM	
Encuesta principal	Encuesta Estructural Empresarial (ENESEM)
Alcance de la investigación	Grandes y medianas empresas Nivel de sección del CIUV4*
Universo de la investigación	12.243 empresas
Técnica de muestreo	Muestreo probabilístico de elementos con selección aleatoria 3.782 empresas 2.561 grandes empresas (Inclusión forzosa) 1.221 medianas empresas (muestreo probabilístico)
Año de referencia	2020
Operativo de levantamiento	Junio - Octubre 2021
Unidad de Análisis	Empresa
Dominios de Investigación	Actividad Económica Principal Tamaño de empresa

Fuente: Módulo Ambiental ENESEM 2020.

La operación estadística denominada **Encuesta Estructural Empresarial (ENESEM)** tiene varios módulos: económico, laboral, ambiental, tecnologías de información y comunicación. En esta guía, mostraremos la información relevante para procesar la información ambiental de las 3777 empresas efectivas con información ambiental válida¹.

En la Figura 1 se muestra el resumen metodológico del MA de la ENESEM 2020, información a conocer previo al uso de la base de datos² y su alcance. A continuación, se hace una breve explicación de la información incluida en este gráfico.

El muestreo de las empresas para la ENESEM 2020 es probabilístico sin repetición, es decir, se escogen sin reemplazo las empresas al azar de un marco maestro referencial de muestreo, el cual ha sido extraído, a su vez, del Directorio Empresarial 2019, el cual es un listado de empresas grandes, medianas y pequeñas que representan al universo de empresas formales del país. Para que los resultados de los cálculos obtenidos a partir de las diferentes variables de la ENESEM sean representativos de las actividades económicas del país, se requiere de la activación de una variable agregada de diseño muestral, denominada "Factor de Expansión", que en el caso de la BDD tiene el nombre "f_exp".

El efecto de aplicar el factor de expansión a la BDD es lograr que todos los resultados de cálculos como agregados estadísticos –promedios, medias, medianas, máximos, mínimos– así como tabulados –como tablas cruzadas o de contingencia–, sean interpretados en términos de todo el marco maestro muestral, del cual puede decirse que tiene una similar estructura o composición de los diferentes tamaños y/o actividades económicas de todo el sector empresarial formal del país. Esto quiere decir, en términos prácticos, que el marco muestral maestro del cual proviene la muestra investigada en la ENESEM 2020 tendría, para la sección económica y laboral, 12243 empresas expandidas (=3782 muestrales), en tanto que para la sección ambiental, son 12238 empresas expandidas (=3777 empresas muestrales).

¹ El recuento de empresas efectivas para los módulos económico y laboral es de 3782. De estas, 5 no tienen información ambiental, por efecto de cobertura parcial de la parte económica. Por tanto, el universo sobre el cual se realizan los cálculos ambientales es de 3777 empresas muestrales, y 12238 empresas expandidas por el factor de expansión.

² La BDD del Módulo de Información Económica Ambiental en Empresas se entrega, a través de la

página web de la operación estadística, en dos versiones: SPSS y CSV. El formato SPSS puede abrirse en cualquier versión SPSS 18 o posterior, mientras que el formato abierto CSV facilita el trabajo de procesamiento de la BDD en plataformas como Microsoft Excel® o en otras más avanzadas, como Stata, R o Python. Sitio web: <https://www.ecuadorencifras.gob.ec/encuesta-de-informacion-ambiental-economica-en-empresas/>, apartado Base de datos.

Sobre el alcance de los datos de la BDD del MA de la ENESEM 2020, o más concretamente, de los cálculos o resultados que se obtienen de ella, en tanto que sean valores absolutos³, serán referenciales (representativos), no de lo que sucede en la economía empresarial nacional, sino únicamente de lo que ocurre en el marco maestro muestral referencial. Sin embargo, como ya se mencionó anteriormente, este marco maestro muestral referencial tiene una composición porcentual –por tamaño y/o actividad económica– de sus empresas que es similar al universo de todas las empresas formales del país –registradas en el SRI–.

Este detalle es muy importante, pues sirve de forma crucial para la interpretación de los resultados que se obtengan de todos los cálculos producidos a partir de las BDD de la ENESEM 2020, concretamente, de la BDD de su Módulo Ambiental. Por ejemplo, si se procesa la variable de consumo de energía eléctrica de red pública (variable **v9001**), y se obtiene la cantidad total de este tipo de energía consumida por las empresas a nivel nacional, y se obtiene el valor de 6,477.24 GWh / año, NO quiere decir que éste fue el consumo real absoluto de las empresas formales del país durante el año 2020. Sin embargo, si en forma similar, se obtiene que el valor de energía eléctrica de red pública consumido por las empresas pertenecientes a la actividad económica "*Industria manufacturera*" es de 3,415.09 GWh / año, entonces Sí se puede afirmar que este sector

aporta al consumo total nacional de energía eléctrica con el $3,619.14 / 6,503.85 = 52.72\%$, tanto para las empresas del marco maestro muestral, como para las empresas del sector formal nacional.

Esto lleva a advertir que, al menos para el Módulo de Información Económica Ambiental en Empresas ENESEM 2020, sus valores absolutos sí son comparables con los valores correspondientes de las mismas variables, pero de años anteriores. Sin embargo, éstos no reflejan el valor real representativo del total de empresas nacionales, a menos que se traten de proporciones (porcentajes), los cuales Sí reflejan, en el mayor grado posible, la realidad de todo el sector formal de las empresas nacionales⁴.

³ Por valores absolutos se entiende todo dato que tiene cantidad y unidad de medida, como 2500 GWh / año de energía. En contraste, los demás valores calculados serán de tipo relativo o porcentual, como puede ser el valor 34.5% de la energía anual consumida.

⁴ Esto se debe a que la muestral de empresas de la ENESEM 2020, derivada del marco maestro muestral, no tiene la misma composición estructural (por actividad económica y/o tamaño de empresa) que la economía de las empresas del país. Cuando se aplica el factor de expansión de

la base de datos, los resultados obtenidos lo son con respecto al marco referencial del marco maestro muestral. Por tanto, las cantidades absolutas son válidas dentro de este marco de referencia, pero no de la economía empresarial nacional. Sin embargo, los valores como proporciones o porcentajes son más estables y se asume cercanos a los valores reales en el marco de la economía empresarial nacional, puesto que el marco maestro muestral tendría una composición estructural similar a la de la economía empresarial formal nacional.

03.

Variables de diseño de la muestra

Variables de diseño de la muestra

Toda muestra probabilística para una encuesta, en particular la de la ENESEM 2020, es una en la cual se debe optimizar la cantidad de empresas pertenecientes a las diferentes categorías de las variables de diseño. En el caso de la ENESEM 2020, son 2 las variables de diseño muestral: (1) El *tamaño de la empresa*⁵; y (2) la *actividad económica* a la cual se dedica. El cruce de c/u de las actividades económicas con c/u de los tamaños se denomina "**dominio**" del universo, o simplemente "dominio muestral".

Esta es la razón por la cual los tabulados e indicadores de publicación se los desagregó, cuando fue posible, por las variables de diseño muestral antes mencionadas. Y esto no es casual: *La incertidumbre de los resultados* –medidos sea con el coeficiente de variación, sea con la longitud del intervalo de confianza al 95% del estimador del parámetro poblacional– se *minimiza cuando se desagregan los resultados por las variables de diseño muestral*. Lógicamente, si el estimador del parámetro poblacional no se desagrega, sino que se calcula solamente a nivel nacional, tendrá el mismo grado de robustez o superior.

Por tanto, no se recomienda desagregar los estimadores poblacionales por otras variables que no sean las de diseño, a menos que el usuario analista de datos descubra alguna(s) variable(s) altamente correlacionadas con las variables de

diseño muestral, y luego desagregue los estimadores poblacionales por estas nuevas variables similares a las de diseño.

Aunque no está contemplado en el alcance de este documento el mostrar la manera de calcular la robustez de los datos, sea a través de medidas de variabilidad como el coeficiente de variación o el intervalo de confianza de los estimadores poblacionales, se menciona al menos que es responsabilidad del usuario analista o calculador de los estimadores el investigar cómo calcular dichas medidas de robustez, pues en función del valor de estas medidas se aprobarán como "estables" o "robustos" los resultados obtenidos, o en su defecto, como valores no estables que se alejan de la realidad del verdadero valor del parámetro poblacional que se estima con los cálculos a partir de la muestra de empresas levantadas cuya información se encuentra en la base de datos de la ENESEM 2020.

⁵ La variable "Ventas netas", junto con la variable "Personal", ambas recabadas en el Directorio Empresarial desde registros administrativos como los del SRI, son las que sirven para definir el tamaño de la empresa. Para mayor información sobre este tema, favor revise el siguiente documento: INEC. (2022). *Metodología de la ENESEM 2020*. INEC, Dirección de Estadísticas Económicas - DECON,

Quito. Recuperado el 11 de abril de 2021, de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Economicas/Encuesta_Estructural_Empresarial/2019/2019_ENESEM_Metodologia.pdf

04.

VARIABLES DE LA BASE DE DATOS DEL MÓDULO DE INFORMACIÓN AMBIENTAL ECONÓMICA EN EMPRESAS Y SU USO

VARIABLES DE LA BDD DEL MÓDULO AMBIENTAL DE LA ENESEM 2020 Y SU USO

Las variables de la BDD del Módulo Ambiental de la ENESEM 2020, atendiendo a su aplicación y utilidad operativa, se clasifican en cuatro tipos:

- (a) *Identificadoras*: Son variables que, como su nombre lo indica, identifican en forma anonimizada a c/u de las unidades de investigación y de análisis, es decir, empresas efectivas de la muestra.
- (b) *De diseño*: Son las variables que se usaron, tanto para calcular el tamaño de la muestra por cada dominio, así como los factores de expansión inicial y final de la BDD.
- (c) *De investigación o temáticas*: Son variables que recaban, según las temáticas señaladas en la parte introductoria, elementos de información de dichas temáticas.
- (d) *Derivadas*: Son variables que resultan de cálculos en cuyas fórmulas intervienen una o más de las variables de investigación y que ameritan ser publicadas en conjunto con la BDD, o bien son variables que separan en categorías útiles para el procesamiento segmentado de datos.

Las *variables identificadoras* de la BDD del MA de la ENESEM 2020, que normalmente son las que encabezan la lista de variables en orden de aparición, son las siguientes:

Tabla 1. Lista de variables identificadoras de la BDD del MA de la ENESEM 2020.

Nombre de la variable	Descripción
inec_identificador_empresa	Identificador de empresa

provincia	Provincia sede de la empresa
cod_tamano	Categoría de tamaño de la empresa
ciiu4_actividad_principal	Código CIIU4 de la actividad económica primaria a 7 dígitos
cod_letra	Código CIIU4 de la actividad económica principal 2020
desc_actividad_principal	Descripción de la Actividad Principal
cod_sector	Categoría del sector económico de la actividad principal de la empresa
ciiu4_actividad_secundaria	Código CIIU4 de la actividad económica secundaria a 7 dígitos
cod_letra_2da	Código CIIU4 de la actividad económica secundaria 2020
desc_actividad_secundaria	Descripción de la Actividad Secundaria
cod_sector2	Categoría del sector económico de la actividad secundaria de la empresa

Fuente: Módulo Ambiental ENESEM 2020.

El uso contemplado para estas variables es el de ordenar la BDD, típicamente en forma ascendente, por el identificador de empresa, el cual es una clave primaria única de identificación anonimizada de cada empresa en la BDD. Las variables *cod_tamano* y *cod_letra*, las cuales pueden ser consideradas también como variables de diseño muestral, sirven para desagregar estimadores poblacionales de todo tipo obtenidos por agregación o filtración de las variables de investigación. Las demás variables identificadoras, excepto *provincia*, sirven para identificar las actividades económicas primaria y secundaria de cada empresa investigada.

Tabla 2. Lista de variables de diseño de la BDD del MA de la ENESEM 2020.

Nombre de la variable	Descripción
N_fin	Tamaño final del dominio

Pst	Categorías de remuestreo de elementos de cada dominio
f_exp	Factor final de expansión de la muestra al marco muestral

Fuente: Módulo Ambiental ENESEM 2020.

Las dos primeras *variables de diseño*, N_fin y Pst, son las que definen el tamaño final de la muestra, así como la migración de dominio de ciertas empresas, todo esto por efecto de la cobertura final de la muestra. En nuestro caso, el MA de la ENESEM 2020 tiene 3,777 empresas efectivas con información ambiental. Sin embargo, al inicio de la fase de levantamiento de datos, se disponían de 4,324 empresas. Con estas empresas, se calculó un factor de expansión inicial, esto es, antes de levantar información de la ENESEM 2020.

Por razones de novedades en la cobertura por falta de información – debido al cierre de la empresa para el año del levantamiento de la información, o a otras causas–, 542 de las 4,324 empresas salieron de la lista final de empresas publicadas. Una parte del resto de empresas efectivas fueron afectadas por novedades varias –como cambio de tamaño y de actividad económica–.

Por todo esto, se necesitó recalcular el factor de expansión, y así se publicó finalmente la variable de expansión f_exp, la cual ya se indicó anteriormente que tiene por objetivo hacer que los resultados de los estimadores poblacionales calculados a partir de la muestra presente en la BDD sean interpretados como representativos del marco muestral maestro de empresas, pero también representativos del sector empresarial formal nacional, sobre todo si los estimadores son de tipo proporción o porcentaje.

Tabla 3. Lista parcial de variables temáticas de la BDD del MA de la ENESEM 2020.

Nombre de la variable	Descripción
v7002	2. Indique el número de personas dedicadas a actividades ambientales
v7003	2.a. Número de personas que trabajaron a tiempo completo
v7004	2.b. Número de personas que trabajaron a tiempo parcial
v7007	3. ¿Cuál fue el costo y gasto en el año 2020 en Protección y Gestión Ambiental? - COSTO AMBIENTAL
v7008	3. ¿Cuál fue el costo y gasto en el año 2020 en Protección y Gestión Ambiental? - GASTO AMBIENTAL
v74	4. Su proyecto, obra o actividad, tiene algún tipo de permiso ambiental (vigente o en trámite) durante el año 2020, Sí / No
v75	5. ¿La empresa contó con la certificación ISO 14001:2015? - Sí / No
v8001	De acuerdo a la información recopilada en el capítulo 4 su empresa invirtió (\$) / año:
v8098	17.TOTAL – PRODUCCIÓN Ambiental: Valor (\$) / año
v8099	17.TOTAL – INVERSIÓN Ambiental: Valor (\$) / año
v8100	17.TOTAL - GASTOS CORRIENTES Ambientales: Valor (\$) / año
v9001	1. Registrar información sobre la Energía Eléctrica comprada a la red pública (área administrativa + productiva) en el año 2020 : - Cantidad / año
v9002	1. Registrar información sobre la Energía Eléctrica comprada a la red pública (área administrativa + productiva) en el año 2020 : - Valor USD
v10000	1. Registrar información sobre agua comprada de Red Pública (área administrativa + productiva) en el año 2020 - Cantidad
v10001	1. Registrar información sobre agua comprada de Red Pública (área administrativa + productiva) en el año 2020 - Valor USD

Fuente: Módulo Ambiental ENESEM 2020.

Las 1,294 *variables de investigación o temáticas* son las que tienen una correspondencia “1 a 1” con cada

pregunta e inciso del instrumento de recolección de información en campo (formulario). A pesar de que todas las variables vienen con sus respectivas etiquetas de variable (descripción) y de valores (categorías, en las variables de este tipo), esto no le exime al usuario/analista de familiarizarse con el formulario de la ENESEM 2020 antes de procesar la BDD del MA de la ENESEM 2020. Además, se recomienda procesar la BDD mencionada con este formulario a la mano. De esta manera, se podrá saber exactamente qué variables ingresar en la sintaxis o en los menús y comandos de la plataforma de procesamiento estadístico para obtener los estimadores poblacionales esperados.

Es importante notar que todas las variables temáticas son de la forma “vXnnn”, donde “X” representa el número del capítulo del MA de la ENESEM 2020 (7 = Gestión ambiental; 8 = Oferta y utilización de bienes y servicios ambientales; 9 = Energía, combustibles y lubricantes; 10 = Agua, manejo de aguas residuales y otros residuos), y “nnn” representa un número entero que va desde “000” y aumenta en forma consecutiva de uno en uno. Así, “v10000” debe ser una variable del capítulo 10, “v7002” una variable del capítulo 7, etc.

Tabla 4. Lista de variables derivadas de la BDD del MA de la ENESEM 2020.

Nombre de la variable	Descripción
valag	Valor Agregado Bruto (VAB) por empresa 2020

⁶ Para mayor información, consultar la página web del Módulo Ambiental de la ENESEM 2020, en sus apartados de: *Tabulados*, donde se puede descargar la versión en Microsoft Excel® de los tabulados e indicadores de la operación estadística, así como el apartado *Sintaxis*, en el cual se puede descargar el archivo de sintaxis en lenguaje de programación SPSS 18+.

Sitio web:

<https://www.ecuadorencifras.gob.ec/encuesta-de-informacion-ambiental-economica-en-empresas/>

W	Impacto ambiental agregado por empresa – Módulo Ambiental ENESEM 2020
---	---

Fuente: Módulo Ambiental ENESEM 2020.

Las dos variables derivadas que incluye la BDD del MA de la ENESEM 2020, son:

- (a) *valag*: Es la variable la cual, calculada por la Dirección de Estadísticas Económicas del INEC a partir de ciertas variables económicas y laborales, define el Valor Agregado Bruto (VAB) por cada empresa. Es una variable continua real valuada.
- (b) *W*: Es un agregado de los logaritmos decimales de las variables escalares –continuas o de valor reales–, así como de otras variables categóricas del MA de la ENESEM 2020, que expresa una medida de impacto ambiental para cada empresa. Es una variable continua en el intervalo numérico [0, 1].

Ambas variables derivadas son utilizadas sobre todo en los indicadores publicados de *intensidad energética*, *intensidad de generación de CO₂* e *intensidad de uso de agua*⁶. También, la variable “W” puede utilizarse para un sinnúmero de análisis de impacto ambiental, económico y mixto, sobre todo para las actividades económicas y para relacionar el impacto ambiental con cualesquiera variables económicas y ambientales⁷. Cabe señalar que, dentro del código del archivo de sintaxis SPSS de indicadores y tabulados, además de

⁷ Se recomienda revisar, en la página web del Módulo Ambiental de la ENESEM 2020, el apartado *Boletín técnico*, en donde se realizan importantes análisis descriptivos que pueden servir de base para realizar análisis más complejos sobre la temática ambiental de las empresas del país. Además, dentro del archivo de sintaxis de tabulados e indicadores, se encuentra al final el código SPSS de construcción del Impacto Ambiental agregado 2020. Se recomienda revisar el código para tener una mejor idea de lo que mide este constructo.

encontrarse la sintaxis del Impacto Ambiental, se encuentra un segmento de código en lenguaje de programación R Markdown 4.1.0, el cual genera una nueva base de datos únicamente de desechos peligrosos⁸.

Esta base de datos no existía en versiones previas de la ENESEM, debido a que los residuos peligrosos se podían procesar directamente a partir de las variables que venían en la BDD del Módulo Ambiental de la ENESEM – en el caso de las variables del año 2020, estas variables están en el rango entre la **v10545** y la **v11006**–. Pero en la BDD del año 2020, el usuario encontrará variables especiales, cuyos nombres son: **tipo_desecho_1**, **tipo_desecho_2**, ..., **tipo_desecho_22**.

Estas variables contienen un texto con un nombre codificado de desecho peligroso, código que pertenece a un catálogo internacional manejado en nuestro país por el Ministerio del Ambiente, y que en la ENESEM 2020 permiten homologar la recolección de datos sobre los mencionados desechos peligrosos catalogados en el país (295 en total). De estos 295 desechos peligrosos, las empresas investigadas en la ENESEM 2020 dieron cuenta de que generan 143 de los 295 desechos peligrosos del catálogo del Ministerio del Ambiente. Cada variable de nombre **tipo_desecho_N** tiene un valor que en general es distinto para cada empresa, por tanto, haciendo que un grupo de las mismas variables contengan información de desechos muy distintos, según el orden en que las empresas hayan escogido llenar la información para los desechos peligrosos que ellas generan.

Así, por ejemplo, en el formulario descargable del Módulo Ambiental

de la ENESEM 2020, puede observarse en la **Línea 384** que el espacio está reservado o sugerido para el ingreso del desecho peligroso de nombre: **“(NE-03) - Aceites minerales usados o gastados”**. mientras que en la **Línea 386**, se ha reservado el espacio para el ingreso del desecho: **“(NE-40) - Luminarias, lámparas, tubos fluorescentes, focos ahorradores usados que contengan mercurio”**. Sin embargo, sucede que en el aplicativo web de la aplicación INFOCAPT de llenado de la información ambiental para la ENESEM 2020, las empresas tuvieron la libertad de invertir el orden sugerido, si ellas lo creían necesario o conveniente. Así, algunas empresas llenaron información sobre el desecho **“(NE-03) - Aceites minerales usados o gastados”** ¡en cualquier línea de ingreso de desechos peligrosos!, y no necesariamente en la **Línea 384**, reservada por diseño del formulario para ese fin. La misma situación ocurrió con buen número de los diferentes desechos peligrosos.

En consecuencia, fue necesario escribir un código relativamente complejo que busque los residuos que tienen el mismo código y nombre, copiar esta información en variables nuevas de una nueva BDD solo de desechos peligrosos, y asignarlas empresa por empresa. La ventaja de esta BDD generada es que está bajo el estándar internacional de la *Forma Normal de Boyce-Codd*⁹, por la cual se garantiza que cada empresa tenga toda la información –de sus desechos peligrosos– en una sola fila, y que cada variable o columna almacene exactamente la misma información para todas las empresas. No obstante, esta BDD no se publica, ya que realmente es una base sintética o generada a partir de la BDD publicada, mediante el código R

⁸ Este código se adjunta como anexo a este documento.

⁹ Sobre el tema de las diferentes normalizaciones de la estructura de bases de datos relacionales, favor ver:

https://es.wikipedia.org/wiki/Normalización_de_bases_de_datos.

Markdown 4.1.0 ya mencionado antes, y que puede ser reconstruida por un usuario / analista en capacidad de hacerlo. En este caso, su procesamiento sería muy similar al de la BDD publicada del Módulo Ambiental de la ENESEM 2020.

La nomenclatura usada para esta BDD generada con los desechos peligrosos por empresa es la siguiente: todo el grupo de variables para un desecho peligroso, por ejemplo el desecho de nombre "**(NE-03) - Aceites minerales usados o gastados**", comenzará con la variable "**y242_0**", la cual contendrá justamente la cadena con el texto "**(NE-03) - Aceites minerales usados o gastados**", para todas las empresas, independientemente si algunas de estas empresas no generan el desecho. Luego, a la derecha de la variable "**y242_0**", aparecerán las variables: "**y242_1**", "**y242_1.1**", "**y242_2.1**", "**y242_2.2**", ..., "**y242_8.1**", "**y242_8.2**", "**y242_8.3**", "**y242_9**", donde el sufijo numérico que viene inmediatamente después del carácter guión bajo "_" es justamente el número de la columna, desde la 1 hasta la 9, que poseen todas y c/u de las tablas de residuos no peligrosos, desechos especiales y desechos peligrosos, pertenecientes al Capítulo 10.III. Así, la variable "**y242_8.2**" es la que almacena el porcentaje del desecho antes indicado que ha sido gestionado por un gestor autorizado. El número "**242**" que aparece a modo de prefijo numérico, inmediatamente después de la letra "**y**", es el secuencial del desecho "**(NE-03) - Aceites minerales usados o gastados**", según aparece en el catálogo de desechos peligrosos homologado internacionalmente por el Ministerio del Ambiente, Agua y Transición Ecológica (MAATE) en nuestro país.

De esta manera, si el usuario /analista de las BDD del Módulo Ambiental de la ENESEM 2020 necesitara procesar la BDD especial generada de desechos

peligrosos, requerirá ejecutar el código ya mencionado, tanto para construir la Tabla 21 perteneciente al libro de Microsoft Excel® publicado con los tabulados e indicadores de la operación estadística, como para realizar otros procesamientos adicionales. El código R Markdown 4.1.0 que genera la base de datos auxiliar (derivada) de desechos peligrosos se lo adjunta a este documento como anexo del mismo.

05.

Recomendaciones generales previas al procesamiento de la base de datos del Módulo de Información Ambiental Económica en Empresas

Recomendaciones generales previas al procesamiento de la BDD del Módulo Ambiental de la ENESEM 2020

Aunque ya se han emitido, en forma algo implícita, un número de recomendaciones importantes previas al procesamiento de la BDD del Módulo Ambiental de la ENESEM 2020, es importante tomar en cuenta estas otras recomendaciones adicionales:

- (1) Iniciar con la lectura de los siguientes documentos que forman parte de la lista de publicables de la operación estadística: Principales resultados, Boletín técnico, tabulados y sintaxis. Se encuentran en el sitio web:
<https://www.ecuadorencifras.gob.ec/encuesta-de-informacion-ambiental-economica-en-empresas/>. De igual manera se recomienda descargarse el formulario y leer el documento de diseño muestral.
- (2) Ejecutar la sintaxis de tabulados e indicadores del Módulo Ambiental de la ENESEM 2020. Esta ejecución tiene que hacérsela por bloques de código. Este código se ha comentado, de manera que cada bloque está separado por comentarios que indican al lector la acción que realiza dicho bloque

de código. Típicamente, cada tabulado e indicador está membretado con los mismos códigos con los que se presenta en el publicable de Tabulados del sitio web de la operación estadística. Así, debería ejecutarse un bloque de código a la vez, y analizar los resultados arrojados por el procesador SPSS.

- (3) Una vez ejecutada la sintaxis SPSS de tabulados e indicadores, se habrán reconstruido estos tabulados e indicadores por parte del usuario / analista. Entonces, podrá modificarse opcionalmente el código para generar otros resultados, sobre todo agregados y desagregados que requiera o necesite el usuario / analista.
- (4) En caso de generarse alguna inquietud acerca del cálculo de algún tabulado o indicador de publicación, así como de cualquier otro resultado en otros publicables, como la presentación de los Principales resultados, o del Boletín técnico, contactarse con la página web del INEC: <https://www.ecuadorencifras.gob.ec/estadisticas/>, e ingresar al menú **Consultas especializadas**, en donde podrá escoger la opción **Requerimiento de Información**. Allí, se encontrará en el Sistema de Atención al Ciudadano, donde podrá realizar cualquier requerimiento de información personalizada, no solamente sobre el Módulo Ambiental de la ENESEM, sino sobre cualquier operación estadística gestionada por el INEC.

06.

**Anexo: Código R
Markdown 4.1.0 para la
generación de la base
de datos derivada de
desechos peligrosos.**

ANEXO: Código R Markdown 4.1.0 para la generación de la BDD de desechos peligrosos

*** IMPORTANTE: El código de generación de la matriz de desechos peligrosos ENESEM 2020, así como el código para su procesamiento, se ha escrito en el lenguaje de programación del entorno de procesamiento estadístico R. El documento fuente es un archivo de tipo "RMarkdown", que es un formato R que permite escribir bloques de código autónomo e interactivo ejecutable, de manera que todos los bloques en conjunto resuelven las tareas ya mencionadas.

- * Código RMarkdown incluido en el archivo "SX Creación BDD Desechos Peligrosos 2020.Rmd".
- * R version 4.1.0 (2021-05-18)
- * RStudio version 1.4.1717 (df86b69e, 2021-05-24) for Windows
- * RMarkdown version 2.9

```
---
title: "Creación BDD Desechos Peligrosos ENESEM 2020"
author: "Ramiro Benavides L."
date: "24/03/2021"
update: "28/04/2022"
output:
  html_notebook: default
  pdf_document: default
---

# 0. Carga de BDD del Módulo Ambiental ENESEM 2020.

```{r}
library(haven)
Carga de la BDD del Módulo Ambiental ENESEM 2020.
B <- read_sav("C:/EMPRESAS 2020/BDD_MA_ENESEM_2020_PUB.sav")
detach("package:haven", unload = TRUE)

Vector de cadenas de texto que contiene los nombres de las variables de identificación que se incluirán
en el frame (o BDD) de reporte de desechos peligrosos.
lista_vars_identif <- c("inec_identificador_empresa")
...

1. Determinación de vector de nombres de desechos peligrosos.

```{r}
Desechos <- NULL

# El archivo "validacion_2019v6.xls", en su hoja "ANEXO_1", contiene el catálogo de los
# 295 desechos peligrosos posibles. Será de utilidad para pasar estos desechos desde la
# BDD del Módulo Ambiental ENESEM 2020 hacia el nuevo frame de desechos peligrosos 2020.
Desechos <- readxl::read_excel("C:/EMPRESAS 2020/validacion_2019v6.xls", sheet = "ANEXO_1")
Desechos[, 2:3] <- NULL
names(Desechos) <- "desecho"
Desechos$desecho[285] <- '(NE-46) - Partes de equipos eléctricos y electrónicos que contienen montajes eléctricos
y electrónicos, componentes o elementos constitutivos como acumuladores y otras baterías, interruptores de
mercurio, vidrios de tubos de rayos catódicos, capacit'
Desechos$var <- NA
for (j in 1:dim(Desechos)[1]) {
  Desechos$var[j] <- paste0("y", j)
}
...

```

2. Creación del dataframe R que contendrá los desechos peligrosos de las empresas.

```
``{r}
R <- B[, lista_vars_identif]
for (j in 1:dim(Desechos)[1]) {
  comando <- paste0("R$y", j, "_0 <- Desechos$desecho[, j, "]")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_1.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_2.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_2.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.1.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.1.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.2.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.2.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.3.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_3.3.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_4.1.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_4.1.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_5 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_6.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_6.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_7.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_7.2 <- as.numeric(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_8.1 <- as.integer(NA)")
  eval(parse(text = comando), envir = .GlobalEnv)

  comando <- paste0("R$y", j, "_8.2 <- as.integer(NA)")
}
```

```
eval(parse(text = comando), envir = .GlobalEnv)
```

```
comando <- paste0("R$y", j, "_8.3 <- as.integer(NA)")  
eval(parse(text = comando), envir = .GlobalEnv)
```

```
comando <- paste0("R$y", j, "_9 <- as.character(NA)")  
eval(parse(text = comando), envir = .GlobalEnv)
```

```
}  
rm(comando, j)  
...
```

3. Copia de desechos peligrosos desde la BDD ENESEM (base "B") al dataframe de desechos peligrosos ("R").

```
```{r}
```

```
Variable que guarda los índices de columna del dataframe B donde aparecen desechos peligrosos.
index <- grep("^tipo", names(B))
```

```
for (i in 1:dim(B)[1]) { # i es índice para las empresas válidas.
```

```
Vector de índices que contiene las posiciones de columna en B de los nombres de desechos peligrosos listos para
copiar al dataframe R.
```

```
sec_OK <- index[which(B[i, index + 1] == 1)]
```

```
if (length(sec_OK) > 0) {
```

```
Vector de cadenas que contiene los nombres de desechos peligrosos a ser copiados de la BDD B hacia R.
```

```
des_OK <- as.character(B[i, sec_OK])
```

```
Vector con los nombres de variables "yi" para llenar en el frame R.
```

```
var_OK <- Desechos$var[which(Desechos$desecho %in% des_OK)]
```

```
Vector que contiene las posiciones de columnas en R de los desechos peligrosos a copiarse.
```

```
ind_R <- as.integer(sapply(des_OK, function(x) {which(R[,] %in% x)}))
```

```
for (j in 1:length(sec_OK)) { # j es índice para desechos peligrosos válidos.
```

```
for (k in 1:21) { # k es índice para las celdas de cada fila de un desecho peligroso.
```

```
temp1 <- B[i, sec_OK[j] + k]
```

```
R[i, ind_R[j] + k] <- temp1
```

```
}
```

```
}
```

```
print(paste0("Iteración = ", i, "; ", "Tiempo del sistema = ", Sys.time()))
```

```
} else {
```

```
next
```

```
}
```

```
}
```

```
rm(temp1, des_OK, i, ind_R, index, j, k, sec_OK, var_OK)
```

```
...
```

# 4. Colocación del valor "2" en las variables de tipo "Ynnn\_1" cuando tienen el valor NA ("missing").

```
```{r}
```

```
indices1 <- grep("\\_1", names(R))
```

```
indices2 <- grep("\\_1.", names(R))
```

```
indices <- setdiff(indices1, indices2) # Este vector guarda los índices de las variables ¿GENERAN (Sí/No)?.
```

```
for (i in 1:dim(R)[1]) {
```

```
R[i, indices] <- ifelse(is.na(R[i, indices]), 2, R[i, indices])
```

```
}
```

```
rm(indices1, indices2)
```

```
...
```

5. OPCIONAL: Obtención de la versión corta de la BDD, esto es, la que contiene únicamente los desechos peligrosos generados por al menos una empresa de publicación.

```
```{r}
```

```

P <- R
for (i in 1:length(indices)) {
 comando <- paste0("a <- min(P$y", i, "_1)")
 eval(parse(text = comando), envir = .GlobalEnv)
 if (a != 1) {
 comando <- paste0("ind_inicial <- which(names(P) == 'y", i, "_0')")
 eval(parse(text = comando), envir = .GlobalEnv)
 P <- P[, -(ind_inicial:(ind_inicial+21))]
 }
}
rm(a, comando, i, ind_inicial)
...

```

# 6. Transformación de todos los desechos peligrosos a kilogramos (se usa la aproximación 1 US gal = 3.78541 kg, en el caso de desechos peligrosos líquidos registrados en galones).

NOTA: SE PODRÁ USAR EL FRAME ESTANDARIZADO "S" (en SPSS) PARA CALCULAR TABULADOS DE DESECHOS PELIGROSOS, UNA VEZ CONCLUIDO ESTE BLOQUE DE CÓDIGO.

```

...{r}
S <- R # S es el frame que contendrá los desechos peligrosos estandarizados a kilogramos.
ind_cantidad <- grep("_0", names(S)) + 4
for (j in ind_cantidad) {
 try(S[which(S[, j-1] == 2), j] <- 1000 * S[which(S[, j-1] == 2), j], silent = TRUE)
 try(S[which(S[, j-1] == 3), j] <- 3.78541 * S[which(S[, j-1] == 3), j], silent = TRUE)
 print(paste0("Columna válida = ", which(ind_cantidad == j), "; ", "Tiempo del sistema = ", Sys.time()))
}
rm(j)
S <- cbind(S, B$f_exp)
names(S)[dim(S)[2]] <- "f_exp"
...

```

# 7. Crear el vector de suma de c/u de las columnas ponderadas de cantidades de desechos peligrosos.

```

...{r}
Z <- S
for (j in ind_cantidad) {
 Z[, j] <- Z[, j] * Z$f_exp # Desechos peligrosos por empresa.
 print(paste0("Columna válida = ", which(ind_cantidad == j), "; ", "Tiempo del sistema = ", Sys.time()))
}
Suma_x_Desecho <- apply(Z[, ind_cantidad], 2, sum, na.rm = T) # Suma de cantidades por desecho peligroso.

```

# Z1 es el frame que reporta la suma de desechos en kg para todos los 156 desechos generados por las empresas, a partir del catálogo de 295 desechos peligrosos.

```

Z1 <- as.data.frame(Suma_x_Desecho) # Frame que contendrá las cantidades ponderadas de los desechos
peligrosos (incluirá el nombre de cada desecho peligroso).
index_names <- which(names(Z) %in% rownames(Z1)) - 4
Z1$nombre_des_pel <- NA
for (i in 1:length(index_names)) {
 Z1$nombre_des_pel[i] <- Z[1, index_names[i]]
}
rm(i, j)
...

```

# 8. Determinar el número de empresas que generan y que conocen las cantidades generadas de desechos peligrosos, por cada desecho peligroso.

```

...{r}
Si_Genera <- apply(Z[, index_names + 1], 2, function(x) {Z$f_exp[which(x == 1)]})
Si_Conoce <- apply(Z[, ind_cantidad], 2, function(x) {Z$f_exp[which(x > 0)]})
Empresas_Genera <- sapply(Si_Genera, sum)
Empresas_Conoce <- sapply(Si_Conoce, sum)
Z1$Genera <- Empresas_Genera

```

```
Z1$Conoce <- Empresas_Conoce
...

```

# 9. Determinar el total de empresas que generan y conocen las cantidades generadas de desechos peligrosos, donde c/u de estas empresas genera o conoce al menos uno de los desechos peligrosos.

```
``{r}
Genera <- apply(Z[, index_names + 1], 1, min, na.rm=T)
Genera[Genera == 2] <- 0
Genera <- Genera * Z$f_exp
Numero_Genera <- sum(Genera)

Conoce <- apply(Z[, ind_cantidad], 1, function(x) {length(which(x > 0))})
Conoce <- as.numeric(as.logical(Conoce))
Conoce <- Conoce * Z$f_exp
Numero_Conoce <- sum(Conoce)
...

```

# 10. Ordenar en forma descendente del vector "Suma\_x\_Desecho". Este procedimiento es útil para determinar la posición 22 del desecho peligroso con la mayor cantidad (masa, kg), ya que la rutina #11 depende de la determinación de este valor. También sirve para determinar los índices de fila en el frame "Desechos", de donde se tomarán los nombres de residuos para adjuntarlos a las cantidades calculadas, en un nuevo dataframe llamado "Resultados2".

```
``{r}
Desechos_ORD_DESC <- sort(Suma_x_Desecho, decreasing = TRUE)
ind <- gsub("_2.2", "", names(Desechos_ORD_DESC))
ind <- gsub("y", "", ind)
ind <- as.integer(ind)
Resultados <- as.data.frame(Desechos_ORD_DESC)
Resultados <- cbind(Desechos$desecho[ind], Resultados)
names(Resultados) <- c("Nombre de desecho", "Cantidad(kg)")
Resultados2 <- Resultados[1:22,]
tempo <- sum(Desechos_ORD_DESC[23:295])
Resultados2 <- rbind(Resultados2, c("Otros residuos", tempo))
...

```

# 11. Determinar el total de empresas que generan y conocen las cantidades generadas de desechos peligrosos, donde c/u de estas empresas genera o conoce al menos uno de los desechos peligrosos, siempre y cuando su cantidad de residuos sea menor que 639485.1. Este valor corresponde a la cantidad expandida del desecho #22, una vez ordenado en forma descendente los valores del vector Suma\_x\_Desecho, el cual tiene las cantidades expandidas (en kg) de desechos peligrosos para todos los desechos con cantidades positivas.

```
``{r}
index_des_menor <- which(Suma_x_Desecho < 639485.1)
Genera_Menor <- apply(Z[, index_names[index_des_menor] + 1], 1, min, na.rm=T)
Genera_Menor[Genera_Menor == 2] <- 0
Genera_Menor <- Genera_Menor * Z$f_exp
Numero_Genera_Menor <- sum(Genera_Menor)

Conoce_Menor <- apply(Z[, ind_cantidad[index_des_menor]], 1, function(x) {length(which(x > 0))})
Conoce_Menor <- as.numeric(as.logical(Conoce_Menor))
Conoce_Menor <- Conoce_Menor * Z$f_exp
Numero_Conoce_Menor <- sum(Conoce_Menor)
...

```

# 12. Exportación de la BDD estandarizada "S" hacia SPSS, así como del frame "Resultados2" a Excel.

```
``{r}
haven::write_sav(S, "C:/EMPRESAS 2020/BDD_Des_Pelig_2020_EST.sav", compress = TRUE)
writexl::write_xlsx(Resultados2, "C:/EMPRESAS 2020/Resultados_DES_PEL_2022.xlsx", col_names = TRUE)
...
#####
FIN DEL CÓDIGO EN RMarkdown de creación y procesamiento de la BDD de desechos peligrosos ENESEM 2020.

```



Elaborado por	Ramiro Benavides	
Revisado por	Carlos Pilataxi	
Aprobado por	Armando Salazar	

**INEC** | Buenas cifras,  
**mejores vidas**



@ecuadorencifras



@ecuadorencifras



@inecEcuador



live/equadorencifras



INEC/Ecuador



INECEcuador



INEC Ecuador