

Guía de uso de Base de Datos del Anuario de Estadísticas de Transporte

Noviembre, 2021

Dirección/Departamento

Dirección de Estadísticas Económicas (DECON)

Unidad

Gestión de Estadísticas Estructurales (GESE)

Elaborado por:

Gabriela Hidalgo
Jorge León

Revisado por:

Roberto Chaves

Aprobado por:

Darío Vélez

Contacto:

inec@inec.gob.ec
www.ecuadorencifras.gob.ec
(02) 2234 164 · (02) 2235 890 · (02) 2526 072

ÍNDICE DE CONTENIDO

Guía de uso de Base de Datos	4
Anuario de Estadísticas de Transporte - ANET 2020	4
1. Introducción	4
2. Datos generales de la operación estadística	4
3. Descripción de las bases de datos	5
4. Cuestionarios o instrumentos de captura	6
5. Anonimización de la Base de datos	7
6. Variables	7
6.1 Variables derivadas	9
7. Tratamiento de valores perdidos, validaciones e imputaciones	10
8. Clasificaciones y Nomenclaturas	11
9. Interpretación de indicadores principales	12
10. Recomendaciones para el usuario	13

ÍNDICE DE TABLAS

Tabla 1 Información del ANET	4
Tabla 2 Descripción de las Bases de Datos del ANET	5
Tabla 3. Lista de principales variables del ANET 2020	7
Tabla 4. Variables derivadas del ANET 2020	9

ÍNDICE DE FIGURAS

Figura 1 Formato de ficha metodológica de indicadores	12
Figura 2 Sección de interpretación del indicador en la ficha metodológica	12

Guía de uso de Base de Datos

Anuario de Estadísticas de Transporte - ANET 2020

1. Introducción

El Anuario de Estadísticas de Transporte es una operación estadística que procesa registros administrativos (RA) para mostrar información relevante sobre: matriculación vehicular, siniestros de tránsito, transporte por ferrocarril, vía aérea y marítimo a nivel nacional; información que es importante a la hora de planificar y organizar los sistemas de transporte del país, por parte de las autoridades competentes y de los diferentes actores públicos y privados.

En este marco, el Anuario presenta información de las siguientes entidades gubernamentales:

- Agencia Nacional de Tránsito (ANT), que proporciona información sobre vehículos motorizados matriculados y siniestros de tránsito;
- Empresa Ferrocarriles del Ecuador Empresa Pública (FEEP), que facilita la información del número de pasajeros movilizados en ferrocarril;
- Dirección General de Aviación Civil (DGAC), que remite la información de tráfico aéreo (Pasajeros y Carga en TM); y,
- Subsecretaría de Puertos y Transporte Marítimo y Fluvial (SPTMF), que entregan información de transporte marítimo y fluvial (Pasajeros y Carga en TM).

El presente documento contiene indicaciones a tener en consideración para el uso adecuado de las Bases de Datos del Anuario de Estadísticas de Transporte 2020.

2. Datos generales de la operación estadística

A continuación, se presenta los datos generales del Anuario de Estadísticas de Transporte 2020:

Tabla 1 Información del ANET

Criterio	Descripción
Nombre de la operación	Anuario de Estadísticas de Transporte
Población Objetivo	<ul style="list-style-type: none"> • Transporte Terrestre: Vehículos motorizados matriculados y Siniestros de tránsito. • Transporte Aéreo: Pasajeros y carga transportada por vía aérea. • Transporte Marítimo: Naves arribadas, pasajeros transportados por vía marítima y carga transportada por vía marítima. • Transporte Ferroviario: Pasajeros transportados por ferrocarril.
Unidad de observación	<ul style="list-style-type: none"> • Vehículos motorizados matriculados: vehículos. • Siniestros de tránsito: personas.

Criterio	Descripción
	<ul style="list-style-type: none"> Transporte aéreo: personas y carga. Transporte marítimo y fluvial: Naves, personas y carga. Transporte ferroviario: personas.
Cobertura geográfica	<ul style="list-style-type: none"> Vehículos motorizados matriculados: Nacional y provincial. Transporte ferroviario: Por filiales, Centro (Quito), Litoral (Durán), Norte (Ibarra), y Sur (Riobamba). Transporte aéreo: Por principales aeropuertos a nivel nacional (Quito, Guayaquil, Esmeraldas, Manta). Transporte marítimo y fluvial: Por puertos públicos nacionales (Esmeraldas, Guayaquil, Manta, Puerto Bolívar, Balao, La Libertad, El Salitral). Siniestros de Tránsito: Nacional, provincial, cantonal.
Tipo de muestreo	N/A
Tamaño de la muestra	N/A
Desagregación de la información	El Anuario de Estadísticas de Transporte ofrece información sobre el número de vehículos matriculados, pasajeros transportados por ferrocarril, carga y pasajeros transportados por vía aérea y marítima, así como, siniestros de tránsito a nivel nacional.
Comparabilidad de la serie histórica	Representación de los resultados en formato tabla, mediante la combinación de variables del último período y datos históricos periodo 2008 - 2020

Fuente: Instituto Nacional de Estadística y Censos

En caso de requerir mayor información acerca del ANET, visite la metodología de la operación estadística en el siguiente enlace:
<https://www.ecuadorencifras.gob.ec/transporte/>

3. Descripción de las bases de datos¹

Las Bases de Datos del ANET 2020 se encuentran disponibles en el siguiente enlace:
<https://www.ecuadorencifras.gob.ec/transporte/>

Tabla 2 Descripción de las Bases de Datos del ANET

Nombre de la Base de Datos	Criterio	Descripción
Vehículos Matriculados	Número de BDD	1
	Formato de la BDD	La base de datos se encuentra en formato EXCEL (.csv) y Spss (.sav).
	Número de registros	2.361.175
	Número de variables (columnas)	10
	Unidad de análisis de la BDD	Vehículos Matriculados
	Bases de datos con las que se puede realizar combinación o cruce de variables (match)	Siniestros de tránsito

¹ No incluye base de Transporte Aéreo, los datos son remitidos a través del Boletín Estadístico de Tráfico Aéreo de la DGAC.

Nombre de la Base de Datos	Criterio	Descripción
Sinistros de Tránsito	Variables necesarias para realizar combinación o cruce de variables (match)	N/A
	Número de BDD	1
	Formato de la BDD	La base de datos se encuentra en formato EXCEL (.csv) y Spss (.sav).
	Número de registros	16.972
	Número de variables (columnas)	11
	Unidad de análisis de la BDD	Sinistros de Tránsito, personas lesionadas y fallecidas.
	Bases de datos con las que se puede realizar combinación o cruce de variables (match)	Vehículos Matriculados
	Variables necesarias para realizar combinación o cruce de variables (match)	N/A
	Número de BDD	1
	Formato de la BDD	La base de datos se encuentra en formato EXCEL (.csv) y Spss (.sav).
Transporte Marítimo	Número de registros	2.582
	Número de variables (columnas)	32
	Unidad de análisis de la BDD	Naves, pasajeros y carga transportada
	Bases de datos con las que se puede realizar combinación o cruce de variables (match)	N/A
	Variables necesarias para realizar combinación o cruce de variables (match)	N/A
	Número de BDD	1
	Formato de la BDD	La base de datos se encuentra en formato EXCEL (.csv) y Spss (.sav).
Transporte Ferroviario	Número de registros	19.887
	Número de variables (columnas)	11
	Unidad de análisis de la BDD	Personas transportadas por ferrocarril
	Bases de datos con las que se puede realizar combinación o cruce de variables (match)	N/A
	Variables necesarias para realizar combinación o cruce de variables (match)	N/A

Fuente: Instituto Nacional de Estadística y Censos

4. Cuestionarios o instrumentos de captura

El ANET es una operación estadística basada en registros administrativos entregados por varias instituciones gubernamentales de acuerdo a sus competencias, como: ANT, DGAC, FEEP, SPTMF.

5. Anonimización de la Base de datos

De acuerdo al principio de confidencialidad estadística o secreto estadístico establecido en el Código de Buenas Prácticas Estadísticas del Ecuador, se garantiza la confidencialidad de la información estadística y resguardo de datos de carácter personal de las unidades de observación.

Las bases de datos que conforman el ANET son sometidas al proceso de anonimización donde se procede con el reemplazo de la variable de identificación directa y eliminación de las variables de identificación auxiliares, tales como: nombres y apellidos, razón social, cédula de identidad, RUC, pasaporte, entre otras. Este proceso permite ocultar la identificación directa del individuo y garantiza la vinculación con otros registros administrativos.

6. Variables

Para el diseño y construcción de las variables, se considera las categorías utilizadas en años anteriores, con la finalidad de estandarizar la información y por ende realizar análisis comparativos y evolutivos de la información procesada, para mayor información puede visitar la **metodología del ANET** en: <https://www.ecuadorencifras.gob.ec/transporte/>

Tabla 3. Lista de principales variables del ANET 2020

Código de la variable	Nombre de la variable	Definición de la variable	Formato del dato de la variable	Categorías de la variable
Vehículos Motorizados Matriculados				
PROVINCIA	Provincia	Indica la provincia de matriculación de los vehículos que existe en el país	Categórica	1 =" Azuay" 2 =" Bolívar" 3 = "Cañar" ... Resto provincias, según DPA.
MODELO	Modelo/ Año de fabricación	Indica el modelo (año de fabricación) de los vehículos matriculados; es decir, es el año en el que se fabricó el vehículo y salió a la venta.	Categórica	1111= "2009 y anteriores" 2010= "2010" 2011= "2011" 2012= "2012" ... 2019="2019" 2020="2020" 2021="2021" ²
MARCA	Marca del vehículo	Determina la marca que predomina en el parque automotor del país.	Cadena	-

² La categoría de esta variable corresponde al año del modelo de vehículo (versiones) que fue matriculado en el 2020.

Código de la variable	Nombre de la variable	Definición de la variable	Formato del dato de la variable	Categorías de la variable
CLASE	Clase del vehículo	Identifica la clase del vehículo que transita en el país.	Categoría	1= "Automóvil", 2= "Autobús", 3= "Camión", 4= "Camioneta", 5= "Furgoneta", 6= "SUV", 7= "Motocicleta", 8= "Tanquero", 9= "Tráiler", 10= "Volqueta", 11= "Otra Clase"
Transporte Ferroviario				
FILIAL	Filial	Indica la filial a la que corresponde el registro de pasajeros	Categoría	1= "Centro" 2= "Litoral" 3= "Norte" 4= "Sur"
RUTA	Ruta	Indica la ruta de los pasajeros.	Cadena	-
MES	Mes de viaje	Indica el mes en que el pasajero viajó	Categoría	1= "Enero" 2= "Febrero" 3= "Marzo" ... 12= "Diciembre"
GEDAD	Grupos de edad	Indica el grupo de edades del pasajero	Categoría	1 = 'De 0 a 11 Años' 2 = 'De 12 a 17 Años' 3 = 'De 18 a 29 Años' 4 = 'De 30 a 64 Años' 5 = 'De 65 y más' 9 = 'No registrado'
SEXO	Sexo del pasajero	Indica el sexo de la persona que se transporta	Categoría	1= 'Hombre' 2= 'Mujer' 9= 'No registrado'
NACIONALIDAD	Nacionalidad del pasajero	Indica la nacionalidad de la persona que se transporta	Categoría	1= 'Nacional' 2= 'Extranjero' 9= 'No registrado'.
CONTINENTE	Continente de origen	Indica el continente de origen de la persona que se transporta	Cadena	-
Transporte Marítimo y Fluvial				
PUERTO	Nombre del puerto	Indica el nombre del puerto en el que las naves entraron y salieron.	Cadena	APE= "Autoridad Portuaria Esmeraldas" APG= "Autoridad Portuaria Guayaquil" APM= "Autoridad Portuaria de Manta" APPB= "Autoridad Portuaria de Puerto Bolívar" SUINBA= "Superintendencia Petrolera de Balao" SUINLI= "Superintendencia Petrolera La Libertad" SUINSA= "Superintendencia Petrolera el Salitral"
TRB	Tonelaje de registro bruto	Indica el tonelaje bruto de las naves entradas y salidas.	Número	-
TRN	Tonelaje de registro neto	Indica el tonelaje neto de las naves entradas y salidas.	Número	-
PASAJ_ENT	Número de pasajeros que entran	Indica los pasajeros entrantes en los diferentes puertos.	Número	-
PASAJ_SAL	Número de pasajeros que salen	Indica los pasajeros salidos en los diferentes puertos.	Número	-
STC	Salida total de carga	Indica la salida total de la carga en toneladas métricas	Número	-
ETC	Entrada total de carga	Indica la entrada total de la carga en toneladas métricas	Número	-
TC	Total de carga	Indica el total de la carga en toneladas métricas	Número	-

Código de la variable	Nombre de la variable	Definición de la variable	Formato del dato de la variable	Categorías de la variable
TRAFICO_C	Tráfico	Indica el tráfico Internacional de las naves	Cadena	-
AGENCIA_NAVIERA_C	Agencia Naviera	Indica el nombre de la agencia naviera con el se transporta la mercancía.	Cadena	-
Siniestros de Tránsito				
PROVINCIA	Provincia	Indica la provincia en la que ocurrió el siniestro de tránsito	Categórica	1 = "Azuay" 2 = "Bolívar" 3 = "Cañar" ... Resto provincias, según DPA.
CLASE	Clases de siniestros de tránsito	Indica las principales clases de siniestros de tránsito en el país.	Categórica	1= "Atropellos" 2= "Caída Pasajeros" 3= "Choques" 4= "Estrellamientos" 5= "Rozamientos" 6= "Volcamientos" 7= "Pérdida De Pista" 8= "Otros "
CAUSA	Causas de siniestros de tránsito	Indica las causas frecuentes de siniestros de tránsito en el país.	Categórica	1= "Embraguez o Droga" 2= "Mal rebasamiento e Invadir carril" 3= "Exceso velocidad" 4= "Impericia e Imprudencia del conductor" 5= "Imprudencia del peatón" 6= "Daños mecánicos" 7= "No respeta las señales de tránsito" 8= "Factores climáticos" 9= "Mal estado de la vía" 10= "Otras Causas"
NUM_LESIONADO	Número de lesionados	Indica el número de lesionados en los siniestros de tránsito.	Numérico	-
NUM_FALLECIDO	Número de fallecidos	Indica el número de fallecidos en los siniestros de tránsito.	Numérico	-
TOTAL_VICTIMAS	Víctimas de siniestros de tránsito	Indica el total de víctimas en los siniestros de tránsito.	Numérico	-

Fuente: Instituto Nacional de Estadística y Censos (INEC)

6.1 Variables derivadas

El ANET construye variables derivadas a partir de la información de la base proporcionada en los registros administrativos de FEED, el MTOP y la ANT; estas derivaciones se realizan con la finalidad de dar una fácil interpretación y comprensión de los resultados y cumplir los objetivos del anuario. Las variables derivadas que se procesan son las siguientes:

Tabla 4. Variables derivadas del ANET 2020

Nº	Nombre de la BDD	Variables derivadas	Descripción
1	2020_VEHICULOS_MATRICULADOS_BDD	ESTRATONE	Es la agrupación simplificada según rangos de capacidad de carga en toneladas que transporta los vehículos motorizados por vía terrestre.
		ESTRAPASAJERO	Comprende la agrupación simplificada según rangos de capacidad de pasajeros que se transportan en vehículos motorizados por vía terrestre.

Nº	Nombre de la BDD	Variables derivadas	Descripción
2	2020_TRANSPORTE_FERROVIARIO_BDD	GEDAD	Es la agrupación por rangos de edad de los pasajeros que se transportan por vía férrea.
		CONTINENTE	Es la agrupación por continente de origen de los pasajeros que se transportaron por vía férrea.
		TCONT	Es la suma de los contenedores con capacidad de carga de 20 y 40 pies.
		STCC	Es la suma de los contenedores de capacidad de carga 20 y 40 pies que registran salidas de carga en el país.
		ETCC	Es la suma de los contenedores de capacidad de carga 20 y 40 pies que registran entradas de carga en el país.
3	2020_TRANSPORTE_MARITIMO_BDD	STC	Es la suma de las salidas del tipo de carga (sólida, líquida, general, contenerizada)
		ETC	Es la suma de las entradas del tipo de carga (sólida, líquida, general, contenerizada)
		TC	Es la suma de las entradas y salidas de carga en los diferentes puertos públicos del país.
		TCONT_20	Es el total de los contenedores tipo 20 (pies) los cuales registran entradas y salidas en los puertos públicos del país.
		TCONT_40	Es el total de los contenedores tipo 40 (pies) los cuales registran entradas y salidas en los puertos públicos del país.
4	2020_SINIESTROS_DE_TRANSITO_BDD	HORA	Son los rangos establecidos en cuanto a la hora del siniestro de tránsito suscitado en el sitio.
		TOTAL_VICTIMAS	Es la suma de personas que se lesionaron y fallecieron en un siniestro de tránsito.

Fuente: Instituto Nacional de Estadística y Censos (INEC)

7. Tratamiento de valores perdidos, validaciones e imputaciones

Validación

Una vez criticada la información de vehículos matriculados, transporte ferroviario, transporte aéreo, transporte marítimo y fluvial y siniestro de tránsito, es sometida a análisis y validación en su totalidad de acuerdo con los siguientes pasos:

- Revisión de consistencia de información
- Derivación de nuevas variables
- Generación y revisión de tabulados preliminares
- Generación y revisión de indicadores
- Generación de archivos preliminares

Imputación

La imputación se considera parte importante del proceso de validación, con el objetivo de presentar un conjunto de datos limpios, completos y coherentes de los RA que conforman esta operación estadística.

El proceso de imputación se aplica específicamente para la sección de Vehículos Matriculados sobre la información proveniente de la ANT, en variables utilizadas para la construcción de los principales resultados, como lo son la capacidad de

carga y pasajeros de los automotores. Para ello se observa los registros que durante la validación y procesamiento reflejaron datos faltantes “missing”, valores en cero y datos atípicos; estos, al ser analizados al detalle por cada una de las clases de vehículo, se presume podría haberse tratado a errores en la digitación, al ingresar la información al sistema de matriculación.

En el análisis de resultados de vehículos según su estratificación de capacidad de carga o de pasajeros, existieron inconsistencias que los usuarios podían interpretar, como una mala validación de los registros, poniendo en duda la calidad de la información.

Para esto se adoptó como método de imputación una variante del modelo “Hot-Deck” el cual tiene como objetivo “llenar los registros vacíos (receptores) con información de campos con información completa (donantes), y los datos faltantes se reemplazan a partir de una selección aleatoria de valores observados, lo cual no introduce sesgos en la varianza del estimador... El algoritmo consiste en ubicar registros completos e incompletos, identificar características comunes de donantes y receptores, y decidir los valores que se utilizarán para imputar los datos omitidos. Para la aplicación del procedimiento es fundamental generar agrupaciones que garanticen que la imputación se llevará a cabo entre observaciones con características comunes, y la selección de los donantes se realiza en forma aleatoria evitando que se introduzcan sesgos en el estimador de la varianza.” (CEPAL; Fernando Medina, Marco Galván, 2007)

En síntesis, para este ejercicio de imputación en la data de matriculación, el modelo utiliza como variables de clasificación para construir los estratos, la “marca” y la “clase” de cada vehículo. Al imputar los datos faltantes, se toman otros de similares características en cuanto a la marca y clase del automotor.

Para realizar el proceso de imputación, se utilizó el software estadístico Stata, ya que facilita la aplicación del modelo, al disponer de un paquete específico para este ejercicio; la sintaxis simula cinco imputaciones diferentes para cada una de las variables tratadas, en este caso capacidad de carga y pasajeros.

Como resultado se obtuvo cifras comparables con respecto a años anteriores, con relación a la clasificación de sus vehículos por capacidad de carga y pasajeros.

8. Clasificaciones y Nomenclaturas

Esta operación estadística utiliza el siguiente Clasificador:

División Política Administrativa (DPA 2020): Es un documento que permite clasificar y homologar la nomenclatura correspondiente a las provincias, cantones y parroquias (urbanas y rurales) del país.

9. Interpretación de indicadores principales

Los principales indicadores del ANET han sido homologados, tomando en cuenta estándares internacionales y criterios expertos. Las fichas metodológicas permiten generar y replicar indicadores con una misma metodología y sirven de insumo al usuario para su análisis³:

Figura 1 Formato de ficha metodológica de indicadores

FICHA METODOLÓGICA	
NOMBRE DEL INDICADOR	Proporción de pasajeros extranjeros que usaron ferrocarril
DEFINICIÓN/OBJETIVO	Conocer la cantidad de pasajeros extranjeros que viajaron por vía férrea en relación al total de extranjeros que ingresaron al país por motivo de turismo.
FÓRMULA DE CÁLCULO	
$PPXF = \left[\frac{NPXTF}{NPXET} \right] \cdot 1000$	
Donde :	
PPXF = Proporción de pasajeros extranjeros transportados por ferrocarril. NPXTF= Número de pasajeros extranjeros transportados por ferrocarril. NPXET= Número de pasajeros extranjeros que entraron al país por cualquier medio de transporte cuyo motivo de viaje fue el turismo.	
DEFINICIÓN DE LAS VARIABLES RELACIONADAS	
Número de pasajeros extranjeros transportados por ferrocarril: Son los pasajeros extranjeros que compraron y utilizaron el boleto para la transportación por ferrocarril por las diferentes rutas del país.	
Número de pasajeros extranjeros entrados por motivo de turismo: Son los pasajeros extranjeros que entraron al país utilizando cualquier medio de transporte, y que además su motivo de viaje haya sido el turismo a las diferentes ciudades del país.	
METODOLOGÍA DE CÁLCULO	
Para obtener la relación de extranjeros que usaron ferrocarril, se considera la información del número de pasajeros extranjeros transportados por ferrocarril de enero a diciembre de 2020, dividido para el número de pasajeros extranjeros entrados por motivo de turismo en el mismo periodo de referencia.	
LIMITACIONES TÉCNICAS	

En el número total de pasajeros extranjeros transportados por ferrocarril puede haber personas que no necesariamente ingresaron al país en el 2020. Por otro lado, el número total de extranjeros que ingresaron al país por motivo de turismo, considera únicamente a los que lo hicieron de enero a diciembre de 2020, y fueron registrados por la autoridad competente.		
UNIDAD DE MEDIDA O EXPRESIÓN DEL INDICADOR	La unidad de medida es: Número de pasajeros extranjeros transportados por ferrocarril.	
INTERPRETACIÓN DEL INDICADOR	Por cada mil extranjeros que ingresaron al país por motivo de turismo, al menos (x) extranjeros se transportaron por vía férrea.	
FUENTE DE DATOS	Instituto Nacional de Estadística y Censos (INEC) – Anuario de Entradas y Salidas Internacionales ESI 2020; Empresa Pública de Ferrocarriles del Ecuador (FEE).	
PERIODICIDAD DEL INDICADOR	Anual.	
DISPONIBILIDAD DE LOS DATOS	2017-2020.	
NIVEL DE DESAGREGACIÓN	GEOGRÁFICO	Nacional.
	GENERAL	No aplica.
	OTROS ÁMBITOS	Mensual.
INFORMACIÓN GEO – REFERENCIADA		No aplica.

Fuente: Instituto Nacional de Estadística y Censos

Las fichas contienen información detallada sobre el indicador, así como su interpretación, por tanto se recomienda revisar el siguiente punto dentro de la ficha del indicador:

Figura 2 Sección de interpretación del indicador en la ficha metodológica

INTERPRETACIÓN DEL INDICADOR	Por cada mil extranjeros que ingresaron al país por motivo de turismo, al menos (x) extranjeros se transportaron por vía férrea.
------------------------------	--

Fuente: Instituto Nacional de Estadística y Censos

Nota: Para mayor información de los indicadores y su forma de cálculo refiérase a los documentos y su sintaxis en el enlace:

<https://www.ecuadorencifras.gob.ec/transporte/>

³ Ficha del indicador en formato establecido por el INEC.

10. Recomendaciones para el usuario

Para un adecuado manejo de la base de datos, se recomienda:

- Revisar la presente guía y los documentos adjuntos como la metodología de la operación estadística.
- Descargar la o las bases de datos en el formato que mejor se ajuste a sus herramientas de trabajo.
- Puede trabajar con el software estadístico de su preferencia.
- Identificar preguntas o indicadores que se desea analizar en la base de datos.
- En caso de tener dificultad del manejo de la base de datos mediante un software estadístico, descargue la base en formato CSV y extraer el archivo de texto a un archivo Excel.



Buenas cifras,
mejores vidas