

Análisis y caracterización de la Base de Datos CENSALUD para el Estudio Epidemiológico del Dengue en Zonas Vulnerables de San Salvador

Sulma Dubon Olmedo

2024-11-01

Tabla de contenidos

Análisis de información y base de datos proporcionado por CENSALUD.	1
Pasos para mejorar la estructura de la tabla	2
Normalización de Datos Faltantes (NULL o NA)	2
Asignar el tipo de dato correspondiente a cada variable en R.	3
Conclusiones	7
Apéndice A: Base de Datos CENSALUD	8
Apéndice B: Tabla de variables, descripción y tipo de datos	9
.	13

Análisis de información y base de datos proporcionado por CENSALUD.

El objetivo es comprender y mejorar la estructura de la tabla de datos proporcionada por CENSALUD para el estudio epidemiológico del dengue en zonas vulnerables de San Salvador y que esta sea adecuada para el análisis en R.

En el [Apéndice A](#), las Figuras [Figura 1](#), [Figura 2](#) y [Figura 3](#) se muestra la base de datos utilizada por CENSALUD. La tabla contiene una estructura compleja para el análisis de datos, debido a la gran cantidad de variables (columnas) empleadas para la recopilación de información. Muchas de estas variables no están definidas claramente, ya que se presentan en forma de preguntas, lo que dificulta su tratamiento.

Además, la base de datos contiene información redundante, algunas preguntas están representadas con columnas separadas para respuestas afirmativas y negativas, categorizadas como “Sí = 1” y “No = 2”. Esta organización, junto con la división en tablas por cada departamento, incrementa la complejidad del análisis.

Pasos para mejorar la estructura de la tabla

Pasos:

- Unificación de datos

La información del estudio se encuentra separada en diferentes tablas de excel por municipio, para unificar los datos se pasan a estructura tabular y se unen las tablas una sola matriz, se incorpora en una columna con el nombre del municipio.

- Eliminación de información irrelevante.

La base de datos cuenta con información personal que no es relevante para el análisis de datos, se eliminaron las variables: Entrevistador, Dirección, Nombre del jefe de familia, Edad y numero de telefono.

- Estandarización y Simplificación de Variables

Las preguntas largas fueron reemplazadas por nombres de variables concisos y descriptivos. En algunos casos, se modificaron el tipo de variable, valores, etc. En el apéndice [Apéndice B](#) se muestra la tabla con sus variables y sus valores

[1] "C:/Users/HP/Desktop/UOC/SEMESTRE3/PRACTICAS/Ptojec_Practicas_UES/Reportes-Q/AnalisisBD"

La tabla con las modificaciones tienen un total de 87 columnas, que son el total de variables a analizar en diferentes áreas, y 21 filas que corresponde al número de muestras tomadas.

Normalización de Datos Faltantes (NULL o NA)

Otro paso importante en el análisis de datos es la a normalización de los valores faltantes, que normalmente conoemos como: NULL, NA, garantizando que todos los valores faltantes estén correctamente identificados.

Resumen de valores NA por columna:

ID	Fecha	Municipio	Coor_Lat	Coor_Long
0	0	0	0	0
CncTrans	CncSint	CncCont	FamDiag	DiagDengue

0	0	0	0	13	
DiagZika	DiagChik	FamHosp	HospDengue	HospZika	
13	13	13	17	17	
HospChik	MetDiag	Caso_6m	AtenMed	PruebLab	
17	13	0	5	5	
AutoMed	MedNat	BrotVec	MedVec	Mosq_uso	
5	5	0	15	0	
Mosq_Insec	Mosq_Peri	MosqTds	MFN<5	MFN5-7	
14	14	14	14	14	
MFA18-40	MFA>40	MosqAguj	MsqFrcStd	Malla_Uso	
14	14	14	14	0	
Repel_Uso	Charla	Fum_Com	Fum_frec	Ult_Vis_A	
0	0	0	6	6	
Fum_Hogar	Fum_MotNo	Fum_Cree	Frec_Limp_Dep	Abat_Uso	
6	20	0	0	0	
Abat_tiem	Inf_prev	Inf_Foll	Inf_char	Inf_tv/rad	
16	0	2	2	2	
Inf_redsoc	inf_otro	Resp_Soc	Resp_Alc	Resp_MINSAL	
2	2	0	0	0	
N<5	N5-17	A18_40	A>40	Embar	
0	0	0	0	0	
ParedTipo	RepelloTipo	TechoTipo	Gote	SueloTipo	
0	0	0	0	0	
EstrucAdic	AguaTipo	FrecRecAg	AlmBarril	AlmCub	
0	0	20	0	0	
AlmPila	AlmOtro	AlmNing	CubTipo	ZancViv	
0	0	0	0	0	
LarvViv	RecBr1	RecCub	RecPila	RecMct	
0	1	1	1	1	
RecLlnts	RecOtros	RecNing	H_Aeg	M_Aeg	
1	1	0	1	1	
H_albo	M_albo				
1	1				

Asignar el tipo de dato correspondiente a cada variable en R.

Un paso muy importante para el análisis de datos es caracterizar cada tipo de datos, esto nos permite utilizar diferentes técnicas para su análisis.

La estructura de la base de datos está completada y lista para su análisis, presentamos una tabla de datos con información esencial para comprender las características de cada variable

Variable	Tipo	Unicos	NA_Cuento	Primeros_Valores
ID	character	21	0	06-17-00-11, 06-17-00-12, 06-17-00-13

Variable	Tipo	Unicos	NA_C conteo Primeros_Valores
Fecha	Date	4	0 2024-08-09, 2024-08-09, 2024-08-09
Municipio	character	2	0 Soyapango, Soyapango, Soyapango
Coor_Lat	character	19	0 13°42.6110'N, 13°42.6120'N, 13°42.3700'N
Coor_Long	character	21	0 89°8.3400'O, 89°8.3390'O, 89°8.2000'O
CncTrans	factor	2	0 1, 1, 1
CncSint	factor	2	0 1, 1, 1
CncCont	factor	2	0 0, 0, 0
FamDiag	factor	2	0 1, 0, 1
DiagDengue	factor	3	13 1, NA, 1
DiagZika	factor	2	13 0, NA, 0
DiagChik	factor	3	13 0, NA, 0
FamHosp	factor	3	13 1, NA, 1
HospDengue	factor	3	17 1, NA, 1
HospZika	factor	2	17 0, NA, 0
HospChik	factor	2	17 0, NA, 0
MetDiag	factor	3	13 1, NA, 1
Caso_6m	factor	2	0 1, 0, 1
AtenMed	factor	3	5 1, 1, 1
PruebLab	factor	3	5 0, 0, 0
AutoMed	factor	3	5 0, 0, 0
MedNat	factor	2	5 0, 0, 0
BrotVec	factor	3	0 1, 1, 1
MedVec	factor	4	15 1, 1, 3
Mosq_uso	factor	2	0 1, 0, 0
Mosq_Insec	factor	2	14 0, NA, NA
Mosq_Peri	factor	3	14 1, NA, NA
MosqTds	factor	2	14 1, NA, NA
MFN<5	factor	2	14 0, NA, NA
MFN5-7	factor	2	14 0, NA, NA

Variable	Tipo	Unicos	NA_Conceo Primeros_Valores
MFA18-40	factor	2	14 0, NA, NA
MFA>40	factor	2	14 0, NA, NA
MosqAguj	factor	3	14 0, NA, NA
MsqFrcStd	factor	3	14 2, NA, NA
Malla_Uso	factor	2	0 0, 0, 0
Repel_Uso	factor	2	0 1, 0, 0
Charla	factor	2	0 1, 1, 0
Fum_Com	factor	2	0 1, 1, 1
Fum_frec	factor	2	6 1, 1, 1
Ult_Vis_A	factor	4	6 3, 3, 3
Fum_Hogar	factor	3	6 1, 1, 1
Fum_MotNo	factor	2	20 NA, NA, NA
Fum_Cree	factor	2	0 0, 0, 0
Frec_Limp_Dep	factor	3	0 1, 3, 3
Abat_Uso	factor	2	0 0, 0, 0
Abat_tiem	factor	3	16 NA, NA, NA
Inf_prev	factor	2	0 1, 1, 1
Inf_Foll	factor	3	2 0, 0, 0
Inf_char	factor	3	2 0, 0, 0
Inf_tv/rad	factor	3	2 0, 1, 0
Inf_redsoc	factor	3	2 1, 1, 1
inf_otro	factor	2	2 0, 0, 0
Resp_Soc	factor	2	0 1, 1, 1
Resp_Alc	factor	2	0 0, 0, 0
Resp_MINSAL	factor	2	0 1, 0, 0
N<5	integer	2	0 1, 0, 0
N5-17	integer	3	0 0, 0, 0
A18_40	integer	4	0 1, 0, 2
A>40	integer	4	0 0, 1, 2

Variable	Tipo	Unicos	NA_Cuento Primeros_Valores
Embar	integer	1	0 0, 0, 0
ParedTipo	factor	1	0 2, 2, 2
RepelloTipo	factor	2	0 1, 4, 1
TechoTipo	factor	3	0 3, 3, 3
Gote	factor	2	0 0, 0, 0
SueloTipo	factor	2	0 3, 3, 3
EstrucAdic	factor	2	0 7, 7, 7
AguaTipo	factor	2	0 1, 1, 1
FrecRecAg	factor	2	20 NA, NA, NA
AlmBarril	factor	2	0 1, 0, 1
AlmCub	factor	2	0 1, 0, 0
AlmPila	factor	2	0 1, 1, 0
AlmOtro	factor	1	0 0, 0, 0
AlmNing	factor	1	0 0, 0, 0
CubTipo	factor	2	0 1, 1, 1
ZancViv	factor	2	0 1, 1, 1
LarvViv	factor	2	0 1, 1, 1
RecBrl	factor	3	1 0, 0, 0
RecCub	factor	3	1 0, 0, 0
RecPila	factor	3	1 1, 0, 1
RecMct	factor	3	1 0, 0, 0
RecLInts	factor	2	1 0, 0, 0
RecOtros	factor	3	1 0, 1, 0
RecNing	numeric	1	0 0, 0, 0
H_Aeg	factor	2	1 1, 1, 1
M_Aeg	factor	3	1 1, 1, 1
H_albo	factor	3	1 0, 0, 0
M_albo	factor	3	1 0, 0, 0

Finalmente guardamos la nueva base de datos con todas las modificaciones en for-

mato.csv

La base de datos normalizada se guardó como 'Encuesta_Normalizada.csv'.

Conclusiones

1. Optimización de la Base de Datos para el Análisis Epidemiológico

La unificación de datos provenientes de múltiples tablas y la eliminación de redundancias permitió transformar una base de datos compleja y fragmentada en una estructura más manejable y adecuada para el análisis en R. Esto facilita la identificación de patrones y tendencias epidemiológicas en las zonas vulnerables de San Salvador.

2. Reducción de la Complejidad en la Organización de Variables

El proceso de estandarización de las preguntas y la simplificación de nombres de variables mejora la legibilidad de la tabla, lo que a su vez reduce errores y simplifica la interpretación. Este paso es clave para realizar análisis más eficientes y menos propensos a confusiones.

3. Eliminación de Información Irrelevante como Práctica Ética y Técnica

La eliminación de datos personales, como nombres, direcciones y números de teléfono, no solo protege la privacidad de los participantes, sino que también reduce el tamaño de la base de datos y elimina ruido innecesario. Esto permite centrarse exclusivamente en las variables de interés para el estudio.

4. Homogeneización de las Tablas por Departamentos y Municipios

La consolidación de tablas separadas en una única estructura tabular con un identificador de municipio simplificó el proceso de análisis multivariable. Este enfoque garantiza que los resultados sean consistentes y permite comparaciones regionales más claras.

5. Contribución al Estudio Epidemiológico del Dengue

Estas mejoras aseguran que los datos recopilados puedan utilizarse de manera más efectiva para identificar factores de riesgo y poblaciones vulnerables. Una base de datos organizada y clara es fundamental para realizar análisis estadísticos confiables y para diseñar estrategias de intervención más precisas.

6. Facilitación del Uso de Herramientas de Análisis en R

La reestructuración de la base de datos se orientó específicamente hacia su compatibilidad con R, lo que permitirá utilizar funciones avanzadas para análisis estadístico, visualización de datos y modelado predictivo. Esto incrementa el potencial del estudio para generar hallazgos útiles y basados en evidencia.

7. Impacto en la Toma de Decisiones para la Salud Pública

Este trabajo no solo mejora la calidad del análisis epidemiológico, sino que también proporciona un modelo de gestión y estructuración de datos que puede replicarse en otros estudios. Esto fortalece la capacidad de CENSALUD para

realizar investigaciones informadas y de alto impacto en el campo de la salud pública.

En conjunto, estas acciones no solo hacen que los datos sean más utilizables, sino que también mejoran la reproducibilidad y la claridad del estudio, lo que permite un análisis epidemiológico más profundo y efectivo del dengue en áreas vulnerables.

Apéndice A: Base de Datos CENSALUD

En esta sección se presentan las figuras que muestran la tabla completa utilizada para el estudio epidemiológico del dengue en zonas vulnerables de San Salvador.

[illegible]

Figura 1: Parte 1 de la base de datos CENSALUD

[illegible]

Figura 2: Parte 2 de la base de datos CENSALUD

[illegible]

Apéndice B: Tabla de variables, descripción y tipo de datos

Variable	Descripcion	Tipo_de_Variable
ID	Codigo vivienda	carácter
Fecha	Fecha encuesta	fecha
Municipio	Municipio	carácter
Coor_Lat	Coordenada latitud	Coordenadas DMS
Coor_Long	Coordenada longitud	Coordenadas DMS
CncTrans	Conocimiento transmisión del dengue	Categorica (Sí=1, No=0)
CncSint	Conocimiento síntomas del dengue	Categorica (Sí=1, No=0)
CncCont	Conocimiento control poblaciones zancudos	Categorica (Sí=1, No=0)
FamDiag	Diagnóstico previo de dengue, zika o chikungunya en la familia	Categorica (Sí=1, No=0)
DiagDengue	Diagnóstico de dengue	Categorica (Sí=1, No=0)
DiagZika	Diagnóstico de zika	Categorica (Sí=1, No=0)
DiagChik	Diagnóstico de chikungunya	Categorica (Sí=1, No=0)
FamHosp	Familia Requiere de hospitalización	Categorica (Sí=1, No=0)
HospDengue	Hospitalización por dengue	Categorica (Sí=1, No=0)
HospZika	Hospitalización por zika	Categorica (Sí=1, No=0)
HospChik	Hospitalización por chikungunya	Categorica (Sí=1, No=0)

Tabla 2: Tabla de Descripción de Variables (*continued*)

Variable	Descripción	Tipo_de_Variable
MetDiag	Método de diagnóstico	Categórica (Prueba de laboratorio=1, Sospechas por sintomatología=2)
Caso_6m	Caso de dengue en los últimos 6 meses	Categórica (Sí=1, No=0)
AtenMed	Atención médica por sospecha de dengue	Categórica (Sí=1, No=0)
PruebLab	Prueba de laboratorio por sospecha de dengue	Categórica (Sí=1, No=0)
AutoMed	Automedicación por sospecha de dengue	Categórica (Sí=1, No=0)
MedNat	Uso de medicina natural por sospecha de dengue	Categórica (Sí=1, No=0)
BrotVec	Conoce brotes de dengue en vecindario últimos 6 meses	Categórica (Sí=1, No=0, Desconoce=2)
MedVec	Medidas tomadas en el vecindario ante brotes de dengue	Categórica (Desconoce=1, Pasar consulta=2, Ninguna=3)
Mosq_uso	Uso de mosquiteros sobre las camas	Categórica (Sí=1, No=0)
Mosq_Insec	Uso de mosquiteros impregnados con insecticida	Categórica (Sí=1, No=0)
Mosq_Perí	Período de uso de mosquiteros	Categórica (Todo el año=1, Época de lluvias=2)
Mosqtds	Uso de mosquiteros en todos los grupos de edad	Categórica (Sí=1, No=0)
MFN<5	Uso de mosquitero familiares menores de 5 años	Categórica (Sí=1, No=0)
MFN5-7	Uso mosquitero familiares entre 5 y 7 años	Categórica (Sí=1, No=0)
MFA1840	Uso mosquitero familiares entre 18 y 40 años	Categórica (Sí=1, No=0)
MFA>40	Uso mosquitero familiares mayores de 40 años	Categórica (Sí=1, No=0)
Mosq_Aguj	Mosquiteros poseen agujeros	Categórica (Sí=1, No=0)
Mosq_frec_est	Frecuencia revisa el estado del mosquitero	Categórica (Seguido=1, Rara vez=2, Nunca=3)
Malla_Uso	Usa mallas en puertas y ventanas	Categórica (Sí=1, No=0)
Repel_Uso	Utiliza algún tipo de repelente	Categórica (Sí=1, No=0)

Tabla 2: Tabla de Descripción de Variables (*continued*)

Variable	Descripción	Tipo_de_Variable
Charla	Ha recibido alguna charla informativa de las autoridades de salud sobre cómo prevenir el dengue	Categórica (Sí=1, No=0)
Fum_Camp	Campañas de fumigación en su comunidad en los últimos 6 meses	Categórica (Sí=1, No=0)
Fum_frec	Frecuencia de las campañas de fumigación en los últimos 6 meses	Categórica (1-2 veces=1, 3-4 veces=2, Más de 4 veces=3)
Ult_Vis_A	Cuando fue la última visita por parte de las autoridades de salud u otras entidades para fumigar su casa	Categórica (1er trimestre=1, 2do trimestre=2, 3er trimestre=3)
Fum_Hogar	Permitió que su hogar sea fumigado	Categórica (Sí=1, No=0)
Fum_MotNo	Motivo por el que no permitió que se fumigara	Categórica (No se encontraba en casa=1, Alergicos en la familia=2)
Fum_Cree	Cree que las fumigaciones son efectivas en reducir la cantidad de zancudos	Categórica (Sí=1, No=0)
Frec_Limp_Dep	Frecuencia de limpieza los depósitos donde se almacena agua en el mes	Categórica (1 vez=1, 2 veces=2, 3 veces=3, más de 4 veces=4)
Abat_Uso	Uso de abate en los depósitos donde almacena agua	Categórica (Sí=1, No=0)
Abat_tiem	Hace cuánto fueron colocadas las bolsas de abate	Categórica (menos de 1 mes=1, alrededor de 2 meses=2, más de 3 meses=3)
Inf_prev	Recibir más información sobre cómo prevenir el dengue	Categórica (Sí=1, No=0)
Inf_Foll	Recibir información en folleto	Categórica (Sí=1, No=0)
Inf_char	Recibir información en charlas comunitarias	Categórica (Sí=1, No=0)
Inf_tv/rad	Recibir información en programas tv/radio	Categórica (Sí=1, No=0)
Inf_redsoc	Recibir información por redes sociales	Categórica (Sí=1, No=0)
inf_otro	Recibir información en otros medios	Categórica (Sí=1, No=0)

Tabla 2: Tabla de Descripción de Variables (*continued*)

Variable	Descripción	Tipo_de_Variable
Resp_Soc	Responsabilidad del control del dengue la sociedad en general	Categorica (Sí=1, No=0)
Resp_Alc	Responsabilidad del control del dengue Alcaldías	Categorica (Sí=1, No=0)
Resp_MINSAL	Responsabilidad del control del dengue Ministerio de salud	Categorica (Sí=1, No=0)
N<5	Niños menores de 5 años	Numérica
N5-17	Niños entre 5 y 17 años	Numérica
A18_40	Adultos entre 18 y 40 años	Numérica
A>40	Adultos mayores de cuarenta	Numérica
Embar	Mujeres embarazadas	Numérica
ParedTipo	Tipo de paredes de la vivienda	Categorica (Adobe=1, Bloque=2, Ladrillo=3, Nahareque=4, Madera=5, Láminas=6, Otro=7)
RepelloTipo	Tipo de repello de la vivienda	Categorica (Cemento=1, Lodo/tierra=2, Parcialmente repellido=3, Sin repello=4)
TechoTipo	Tipo de techo de la vivienda	Categorica (Palma=1, Paja=2, Lámina=3, Duralita=4, Zinc=5, Madera=6, Tejado=7, Otro=8)
Gote	Presencia de goteras en el techo	Categorica (Sí=1, No=0)
SueloTipo	Tipo de piso de la vivienda	Categorica (Tierra=1, Cemento=2, Ladrillo=3, Otro=4)
EstrucAdic	Estructuras adicionales en la vivienda	Categorica (Gallinero=1, Porqueriza=2, Cocina de leña/horno=3, Conejera=4, Trapiche=5, Otro=6, Ninguno=7)
AguaTipo	Tipo de agua utilizada en casa	Categorica (ANDA=1, De pozo=2, Río=3, Nacimiento=4, Pipa de agua=5, Otro=6)
FrecRecAg	Frecuencia de recolección de agua	Categorica (1 vez=1, 2 veces=2, 3 o más veces=3)
AlmBarril	Almacena en barriles	Categorica (Sí=1, No=0)
AlmCub	Almacena en cubetas	Categorica (Sí=1, No=0)
AlmPila	Almacena en pila	Categorica (Sí=1, No=0)
AlmOtro	Almacena en otro	Categorica (Sí=1, No=0)
AlmNing	No almacena	Categorica (Sí=1, No=0)

Tabla 2: Tabla de Descripción de Variables (*continued*)

Variable	Descripcion	Tipo_de_Variable
CubTipo	Tipo de cubierta en los recipientes de almacenamiento de agua	Categorica (Ninguno=1, Tapaderas=2, Pedazos de plástico=3, Láminas=4, Otro=5)
ZancViv	Presencia de zancudos adultos en la vivienda	Categorica (Sí=1, No=0)
LarvViv	Presencia de larvas en la vivienda	Categorica (Sí=1, No=0)
RecBrl	Recolecta larvas en barril	Categorica (Sí=1, No=0)
RecoCub	Recolecta larvas en cubetas	Categorica (Sí=1, No=0)
RecPil	Recolecta larvas en pila	Categorica (Sí=1, No=0)
RecMct	Recolecta larvas en macetas	Categorica (Sí=1, No=0)
RecLlnts	Recolecta larvas en llantas	Categorica (Sí=1, No=0)
RecOtr	Recolecta larvas en otros	Categorica (Sí=1, No=0)
H_Aeg	Presencia de hembra aegypti en la vivienda	Categorica (Sí=1, No=0)
M_Aeg	Presencia de macho aegypti en la vivienda	Categorica (Sí=1, No=0)
H_albo	Presencia de hembra Albopictus	Categorica (Sí=1, No=0)
M_albo	Presencia de macho Albopictus	Categorica (Sí=1, No=0)