



# RECODIFICACIÓN DE VARIABLES Y CRUCE DE BASES DE DATOS

**Diana Maria Perilla Galvis**

Estadística - Secretaría de Salud Departamental



**GOBERNACIÓN**  
Departamento del  
**Valle del Cauca**



[www.valledelcauca.gov.co/](http://www.valledelcauca.gov.co/)





# RECODIFICACIÓN DE VARIABLES Y CRUCE DE BASES DE DATOS

- Diana Maria Perilla Galvis
- Estadística - Secretaría de Salud Departamental

# INTRODUCCIÓN



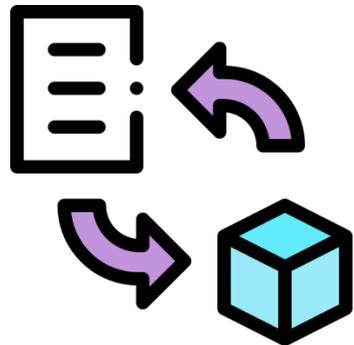
La gestión de datos es el pilar de la investigación en salud. En todo estudio cuantitativo, tener variables bien definidas y datos integrados permite que los análisis sean más robustos y las conclusiones más confiables. La recodificación de variables y la correcta fusión de bases son prácticas esenciales para adaptar la información a los requisitos analíticos, facilitar la interpretación y fortalecer la comparación entre investigaciones.



# RECODIFICACIÓN DE VARIABLES



La recodificación de variables se refiere a transformar o reasignar valores de variables existentes, o convertir variables continuas en categorías, de acuerdo con el contexto y objetivos del estudio (Feng et al., 2020).



Este proceso ayuda a que los datos se ajusten mejor a los supuestos estadísticos de las pruebas utilizadas (como la normalidad), mejora la claridad al presentar resultados y permite estandarizar la información entre estudios o entre bases de datos distintas (Osborne, 2002).

# RECODIFICACIÓN DE VARIABLES



## Razones para la recodificación

### Simplificar variables

Agrupar múltiples categorías en menos categorías para facilitar el análisis e interpretación.

### Crear Nuevas Variables

Convertir una variable continua o discreta en grupos o rangos.

### Adaptar Datos a Modelos

Preparar los datos para que se ajusten a ciertos modelos estadísticos.

### Análisis Exploratorio

Identificar patrones o relaciones en los datos.

# RECODIFICACIÓN DE VARIABLES



## Tipos de recodificación

### Recodificación por rangos:

Se agrupan  
valores continuos  
en intervalos  
definidos

### Recodificación simple:

Cambio directo de  
códigos

### Recodificación condicional:

Basada en varios  
criterios

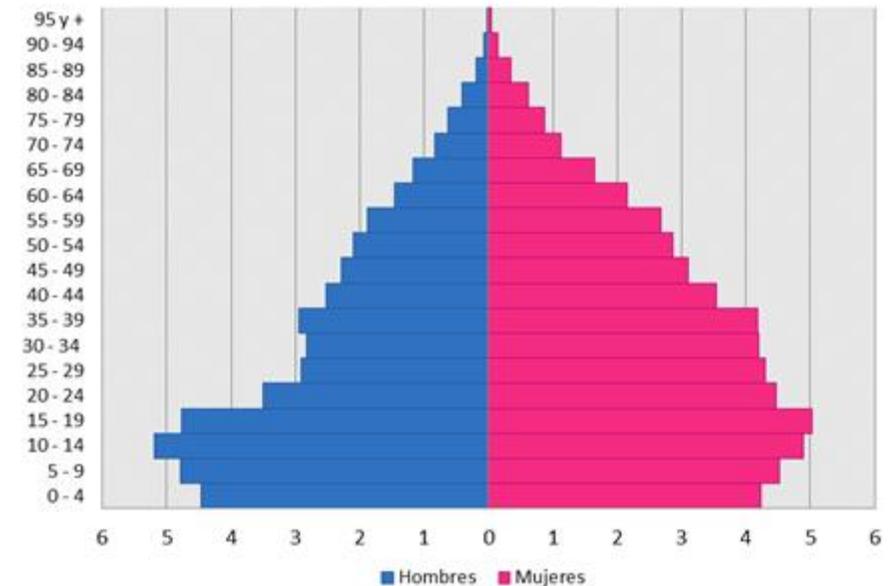
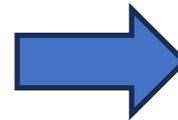
# RECODIFICACIÓN DE VARIABLES



## Ejemplos

Edad simple  
a edad por  
quinquenios

edad	edad_agrup
40	40 a 44
1	0 a 4
2	0 a 4
26	25 a 29
39	35 a 39
28	25 a 29
15	15 a 19
7	5 a 9
7	5 a 9
6	5 a 9
25	25 a 29
6	5 a 9
28	25 a 29
10	10 a 14
27	25 a 29



# RECODIFICACIÓN DE VARIABLES



## Ejemplos

**Porcentaje de un indicador a Nivel de desempeño**

porcentaje_ind	nivel_desempeño
45	Bajo desempeño
93	Desempeño medio alto
82	Desempeño medio bajo
80	Desempeño medio bajo
100	Alto desempeño
90	Desempeño medio alto
50	Bajo desempeño
90	Desempeño medio alto
39	Bajo desempeño
56	Bajo desempeño
67	Bajo desempeño
100	Alto desempeño
96	Alto desempeño

Escala de evaluación	Porcentaje
Bajo desempeño	0 - 70
Desempeño medio bajo	71 - 89
Desempeño medio alto	90 - 94
Alto desempeño	95 - 100

**Recodificación con varios criterios:  
Tres condiciones simultáneas**

tiene_hijos	estado_civil	edad	condicion (madre soltera menor de edad)
si	soltera	17	cumple
si	casada	22	no cumple
no	soltera	20	no cumple
si	soltera	15	cumple
no	casada	31	no cumple
no	separada	23	no cumple
si	union libre	24	no cumple
si	soltera	16	cumple
si	casada	31	no cumple
no	casada	19	no cumple

# RECODIFICACIÓN DE VARIABLES



## Recomendaciones generales

### Crear una nueva variable

- No sobrescribir la variable original para no perder la información inicial y tener un respaldo.

### Usar etiquetas para la nueva variable

- Asignar etiquetas a los valores de la nueva variable para entender claramente qué significa cada categoría recodificada.

### Verificar los resultados

- Revisar las frecuencias de la nueva variable para asegurarse de que la recodificación se realizó correctamente.

# RECODIFICACIÓN DE VARIABLES



## Función SI

La función SI permite realizar comparaciones lógicas entre un valor y un resultado esperado.

- **Estructura básica**

=SI(**condición**; **valor\_si\_verdadero**; **valor\_si\_falso**)

Variable edad ➡ =SI(A2>=18; “Adulto”; “Menor”)

Variable sexo ➡ =SI(B2>=“F”; “Femenino”; “Masculino”)

# RECODIFICACIÓN DE VARIABLES



## Función SI anidada

- ❑ Es cuando se coloca una función SI dentro de otra, para evaluar más de dos categorías o condiciones de forma secuencial.
- ❑ La estructura general de una función SI anidada implica colocar una o más funciones SI completas dentro del argumento "valor\_si\_falso" de otra función SI, formando una cadena de pruebas lógicas.

```
=SI(prueba_lógica1; valor_si_verdadero1; SI(prueba_lógica2;  
valor_si_verdadero2; SI(prueba_lógica3; valor_si_verdadero3; ...)))
```

# RECODIFICACIÓN DE VARIABLES



## Función SI anidada

### Cómo funciona

- La primera función SI: evalúa la "prueba\_lógica1".
- Si la prueba\_lógica1 es verdadera, se devuelve "valor\_si\_verdadero1".
- Si la prueba\_lógica1 es falsa, la función SI principal pasa a su argumento "valor\_si\_falso", que contiene una nueva función SI.
- Esta segunda función SI: (y las siguientes) evalúa "prueba\_lógica2", y así sucesivamente, creando una serie de evaluaciones que se ejecutan secuencialmente si las condiciones anteriores fueron falsas.

### Ejemplo de estructura

- Si la calificación es  $> 90$ , entonces "Excelente".
- Si no, pero la calificación es  $> 70$ , entonces "Buena".
- Si no, pero la calificación es  $> 50$ , entonces "Regular".
- Si no, entonces "Deficiente".

=SI(A2>90; "Excelente"; SI(A2>70; "Buena"; SI(A2>50; "Regular"; "Deficiente"))))

# RECODIFICACIÓN DE VARIABLES



## Función SI.CONJUNTO

- ❑ Permite evaluar múltiples condiciones sin necesidad de anidar varias funciones SI.
- ❑ Es más ordenada, legible y fácil de manejar, especialmente cuando se tienen muchas categorías.
- ❑ Evalúa las condiciones en orden.
- ❑ Devuelve el valor correspondiente a la primera condición que se cumpla.
- ❑ Si ninguna condición se cumple, puede devolver error

```
=SI.CONJUNTO(condición1; valor1; condición2; valor2; ...; condiciónN,  
valorN)
```

# RECODIFICACIÓN DE VARIABLES



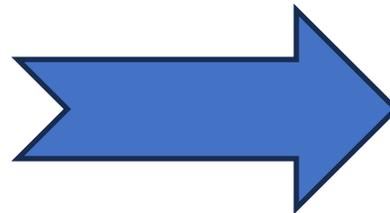
## Función SI.CONJUNTO

Si ninguna condición se cumple, puede devolver error.

=SI.CONJUNTO(A2<10;"primer gr";A2<20;"segundo gr";A2<30;"tercer gr")

puntaje	grupo
11	segundo gr
25	tercer gr
7	primer gr
80	#N/D
96	#N/D
45	#N/D
23	tercer gr
36	#N/D
4	primer gr
50	#N/D
39	#N/D
0	primer gr

Solución



=SI.CONJUNTO(A2<10;"primer gr";A2<20;"segundo gr";A2<30;"tercer gr";VERDADERO;"cuarto gr")

puntaje	grupo
11	segundo gr
25	tercer gr
7	primer gr
80	cuarto gr
96	cuarto gr
45	cuarto gr
23	tercer gr
36	cuarto gr
4	primer gr
50	cuarto gr
39	cuarto gr
0	primer gr

# RECODIFICACIÓN DE VARIABLES



## Combinando función SI y función Y

Estructura de la función Y

=Y(condición1; condición2; ...) ➡ =Y(A2>25; B2="Femenino")

Devuelve VERDADERO solo si ambas condiciones se cumplen:  
edad > 25 y sexo femenino

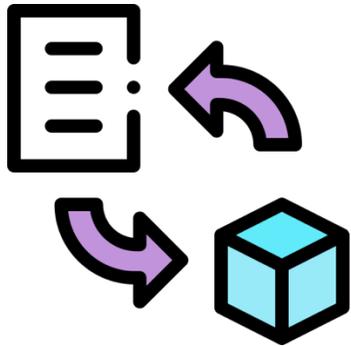
Ejemplo funciones combinadas: Asignar "Grupo A" si la persona es **mayor de 30** y tiene **escolaridad universitaria**

=SI(Y(A2>30; C2="Universitaria"); "Grupo A"; "Otro")

# RECODIFICACIÓN DE VARIABLES



## Recomendaciones

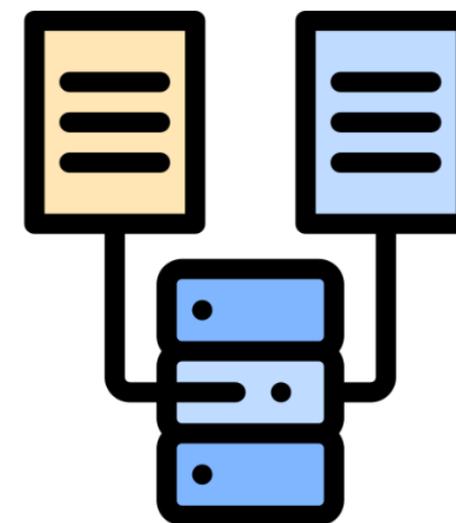


- ❑ Las condiciones de **texto** (como sexo o escolaridad) deben ir entre **comillas**
- ❑ Comprender bien la lógica condicional
- ❑ Evitar anidamientos innecesarios
- ❑ Cerrar bien los paréntesis
- ❑ Probar con ejemplos pequeños antes de aplicar masivamente

# CRUCE DE BASES DE DATOS



- El cruce de bases de datos es el proceso de combinar y comparar información de diferentes conjuntos de datos para identificar patrones, depurar errores, centralizar datos y obtener un análisis más completo.
- Esto se logra buscando y enlazando registros comunes entre las bases de datos mediante columnas o claves compartidas

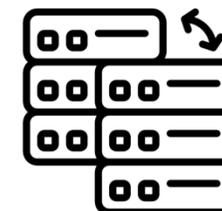


# CRUCE DE BASES DE DATOS



- Función **BUSCARV**

Permite relacionar tablas de datos



La misión de la función **BUSCARV** de Excel es buscar un determinado dato en la primera columna de una tabla o matriz y una vez localizada la fila en la que se encuentra dicho dato, devolver el valor que tiene en esa misma fila la columna que especifiquemos.

Estructura

=BUSCARV (valor\_buscado; matriz\_tabla; indicador\_columnas; [rango])

# CRUCE DE BASES DE DATOS



Número Id	CRUCE 1	Edad	Unidad de Medida
1112106893	=BUSCARV(	27	AÑOS
1117018049	BUSCARV(valor_buscado; matriz_tabla; indicador_columnas; [rango])		
14450323		72	AÑOS

- **Valor buscado:** valor que desea buscar, también conocido como el valor de búsqueda
- **Matriz buscar en:** rango donde se encuentra el valor de búsqueda. El valor de búsqueda debe estar siempre en la primera columna del rango para que BUSCARV funcione correctamente. Por ejemplo, si el valor de la búsqueda está en la celda C2, su rango debería empezar con C
- **Indicador columnas:** número de columna del rango que contiene el valor devuelto
- **Ordenado:** corresponde a la coincidencia de la operación. Si es *VERDADERA* (1) corresponde a una coincidencia aproximada o si es *FALSA* (0) corresponde con una coincidencia exacta.

# CRUCE DE BASES DE DATOS



## Aspectos Clave y Recomendaciones

- ❑ **Posición del valor de búsqueda:**  
La columna donde se encuentra el valor de búsqueda siempre debe ser la primera columna de la matriz (rango) especificada.
- ❑ **Número de columna de retorno:**  
Contar desde la izquierda, la primera columna de tu rango como 1, la segunda como 2, y así sucesivamente, para indicar qué columna debe devolver el valor.

num_ide_	nmun_resi	cod_ase_	nom_eve
1006464971	CAICEDONIA	ESS118	AGRESIONES POR ANIMALES POTENC
1113778010	BOLIVAR	EPS041	VIGILANCIA EN SALUD PÚBLICA DE LA
1112792543	CARTAGO	ESSC24	VIGILANCIA EN SALUD PÚBLICA DE LA
1105384177	PALMIRA	EPS018	DENGUE
66767185	PALMIRA	EPS018	DENGUE
1113673458	PALMIRA	EPS018	DENGUE
1107847959	PALMIRA	EPS018	DENGUE
1006108648	CALI	ESS118	VIH/SIDA/MORTALIDAD POR SIDA
1116236219	TULUA	EPS005	CÁNCER DE LA MAMA Y CUELLO UTER
29613500	TULUA	EPS005	CÁNCER DE LA MAMA Y CUELLO UTER
3118979	TULUA	ESS118	CÁNCER DE LA MAMA Y CUELLO UTER
1130591014	CALI	ESS118	VIH/SIDA/MORTALIDAD POR SIDA
29121928	CALI	ESS024	VIH/SIDA/MORTALIDAD POR SIDA
1130649627	CALI	ESSC18	VIH/SIDA/MORTALIDAD POR SIDA
34505362	CALI	EPS037	TUBERCULOSIS
29502147	FLORIDA	ESS118	AGRESIONES POR ANIMALES POTENC
1006466281	TULUA	EPS018	ENFERMEDAD TRANSMITIDA POR ALI

# CRUCE DE BASES DE DATOS



## Aspectos Clave y Consejos

### Coincidencia exacta o aproximada:

El último argumento de la función (ordenado) determina si la búsqueda es exacta o aproximada.

FALSO: Para una coincidencia exacta del valor buscado, utiliza FALSO o 0. Este es el escenario más común.

VERDADERO: Para una coincidencia aproximada, utiliza VERDADERO o 1. Sin embargo, para que funcione correctamente, la primera columna de tu rango debe estar ordenada ascendentemente.

### Búsqueda de texto y números:

BUSCARV puede buscar tanto valores de texto como valores numéricos.

### No distingue mayúsculas y minúsculas:

La función realiza la búsqueda de manera que "hola" y "Hola" se consideran el mismo valor.

### Dependencia de la primera columna:

La búsqueda siempre se realiza de izquierda a derecha

# CONCLUSIONES



- La recodificación transforma datos para optimizar qué y cómo analizamos, garantizando que la información refleje categorías y rangos relevantes para el contexto médico y epidemiológico.
- La fusión de bases de datos, por su parte, multiplica la riqueza del análisis permitiendo la integración de distintas dimensiones y poblaciones bajo estudio. Ambas técnicas requieren un enfoque estratégico, ético y bien documentado como parte esencial de la calidad científica.





# GRACIAS





Elaborado por:

**Diana Maria Perilla Galvis**

The Global Health Network LAC – Universidad del Valle

Este material de aprendizaje electrónico es propiedad de la Red Global de Salud. Usted es libre de compartir o adaptar este material, pero debe atribuirlo a The Global Health Network utilizando el enlace <https://lac.tghn.org/>

Mayor información en: [proyecto.globalhealthnetworklac@correounivalle.edu.co](mailto:proyecto.globalhealthnetworklac@correounivalle.edu.co)