



Cómo construir una base de datos

How to build up a database

Seung Hyun Jeong, *, ‡ Liliana Paola Farfán Lara, *, § José María Jiménez Ávila, ¶



Palabras clave:
base de datos, investigación,
variable, estadística.

Keywords:
database, research,
variable, statistics.

RESUMEN

En los últimos años ha habido grandes avances y cambios en investigación en el área de la medicina y gran parte de estos se debe al uso de bases de datos. Un adecuado uso de esta herramienta requiere gran destreza de parte del investigador, quien debe saber cómo definir una pregunta de investigación a partir de la cual se determinarán las variables que van a ser estudiadas a través del trabajo de investigación. Aunado a lo anterior, el investigador debe conocer cómo recopilar los hallazgos que surgen a partir de estas variables y la manera más adecuada de analizar y reportar los resultados de dicho estudio. El objetivo principal de cualquier trabajo de investigación es generar información en forma de datos con la finalidad de responder una pregunta de investigación. La cantidad de información que es estudiada en un trabajo de investigación puede ser muy extensa. Para hacer más eficiente este proceso, se recomienda el uso de las bases de datos para facilitar el análisis de la información y a su vez garantizar la validez y reproducibilidad del estudio. Para cumplir con lo anterior, como paso inicial es importante tener claro el objetivo que se desea alcanzar mediante el uso de la base de datos, ya que esto ayuda a minimizar los sesgos y errores durante el proceso de recopilación de la información. Las bases de datos facilitan la colaboración entre diferentes investigadores, quienes pueden acceder a la misma información y trabajar de manera conjunta desde distintas perspectivas. Esto promueve un enfoque más interdisciplinario y enriquecedor en la investigación científica, ampliando todavía más los alcances del estudio. El desarrollo de tecnologías informáticas ha permitido la creación de bases de datos cada vez más estructuradas, que permiten al mismo tiempo la búsqueda, recuperación y análisis de datos. Estas herramientas son indispensables en la investigación hoy en día, ya que permiten manejar y analizar grandes volúmenes de información de manera eficiente.

ABSTRACT

In recent years, there have been significant advances and changes in research in the medical field, mainly due to the use of databases. Proper use of this tool requires great skill from the researcher, who must know how to define a research question from which the variables to be studied through the research work will be determined. In addition to the above, the researcher must know how to gather the results that arise from these variables and the most appropriate way to analyze and report these results. The main objective of any research is to generate information in the form of data in order to answer a research question. The amount of information studied in a research can be very extensive. In order to make this process more efficient for the researcher, it is recommended to use databases to facilitate information analysis and, at the same time, ensure the validity and reproducibility of the study. To achieve this, it is important to have a clear objective for using the database as an initial step, as this helps minimize biases and errors during the information gathering process. Databases facilitate the collaboration among different researchers, who can access the same information from different perspectives. This promotes a more interdisciplinary and enriching approach to scientific research, further expanding the scope of the study. The development of technologies allowed the creation of increasingly structured databases that enable the search, retrieval, and analysis of data. These tools are indispensable in research nowadays, as they allow for the efficient management and analysis of large volumes of information.

* Instituto Tecnológico y de Estudios Superiores de Monterrey, campus Guadalajara. Guadalajara, Jalisco, México.

† ORCID: 0000-0001-8586-5246

‡ ORCID: 0009-0002-0919-9450

¶ Centro Médico Nacional de Columna. Instituto Mexicano del Seguro Social. ORCID: 0000-0002-5532-5318

Correspondencia:
Seung Hyun Jeong
E-mail: seunghyun217@gmail.com

Citar como: Hyun JS, Farfán LLP, Jiménez ÁJM. Cómo construir una base de datos. Cir Columna. 2024; 2 (3): 194-198. <https://dx.doi.org/10.35366/116924>



INTRODUCCIÓN

Las bases de datos en el área de la medicina permiten la recopilación y análisis de grandes cantidades de información de manera detallada, mejorando la calidad del trabajo de investigación, además, el uso de bases de datos facilita la identificación de patrones y tendencias de esta gran cantidad de información, reforzando la validez de los hallazgos del estudio. Teniendo en cuenta lo anterior, es importante aprender acerca de los elementos que componen una base de datos para conocer el adecuado uso de este. Las variables son elementos esenciales de cualquier trabajo de investigación científica, a través de las cuales el investigador formula una hipótesis y extrae sus conclusiones, una vez identificadas las variables que se pretenden estudiar, estas deben ser sometidas a un proceso conocido como “operacionalización de las variables”, a través del cual se realiza una definición precisa de estas variables para que puedan ser medidas para su posterior análisis. Una vez recopilados los datos de interés para el estudio, el investigador debe encargarse de realizar la representación gráfica de los hallazgos para su análisis estadístico.

En este artículo se presentan los tipos de variables utilizados en un trabajo de investigación, el proceso de la operacionalización y las diferentes modalidades de recolección de información para una adecuada estructuración de una base de datos científica.

Variables en estadística

Una variable es cualquier elemento que se puede medir u observar y que puede ser registrada en un conjunto de datos.¹ Las variables pueden tomar diversos valores, desde características de un fenómeno o evento, hasta la persona o grupos dentro de una población que participan en el estudio. Estos valores pueden sufrir cambios entre una situación y otra dependiendo de la intervención que se pretende estudiar, por lo que estos pueden representar diferentes valores. Como cualquier componente de un estudio de investigación, las variables deben ser estructuradas de tal manera que sean objetivas y así reducir el riesgo de sesgos. Las variables deben ser precisas, es decir, cada vez que se midan deben presentar la mínima variación posible, otra característica a destacar es la exactitud, la cual determina qué tan cercano se encuentra la variable que se está estudiando al estándar de referencia, las variables se pueden clasificar de diferentes maneras dependiendo de su naturaleza.² En

la *Tabla 1* se resume la clasificación de las variables junto con ejemplos de su aplicación en un trabajo de investigación.¹⁻⁷

Aplicación de la estadística en los datos

La decisión del tipo de análisis estadístico que se va a utilizar depende de la naturaleza de la variable que se va a estudiar. Desde un principio el diseño de investigación debe incluir las técnicas estadísticas que se planean utilizar, esto con la finalidad de evitar el uso de pruebas estadísticas inadecuadas para el diseño de investigación por el hecho de ser el más avanzado o el único conocido por el investigador, lo cual suele ser un error frecuente en la aplicación del análisis estadístico. Los puntos que se deben tomar en cuenta para decidir el tipo de prueba estadística a utilizar son: las características de las variables, la pregunta de investigación, la hipótesis planteada, la potencia y eficiencia de la prueba elegida y las características de la muestra que se estudiará.⁸

Estadística descriptiva: la estadística descriptiva se encarga de la recolección y síntesis de una serie de datos, donde el enfoque principal es “describir” los datos que se presentan en el estudio, sin realizar una deducción de los fenómenos que se están analizando, esto mediante el uso de medidas de tendencia central y de dispersión.^{7,9} Las medidas de tendencia central, las cuales incluyen la media aritmética, mediana y la moda, describen la distribución de un grupo de datos.

Estadística inferencial o analítica: la estadística inferencial es aquella que “analiza” una serie de datos con la finalidad de llegar a una conclusión acerca del estudio, para esto, es necesario formular una hipótesis y con base en ellos, determinar si los hallazgos del estudio concuerdan con esta hipótesis.⁹

Estadística paramétrica y no paramétrica: las pruebas estadísticas paramétricas se utilizan en variables que siguen una distribución y varianza equitativa entre ellas. Los métodos no paramétricos no toman en cuenta la distribución de las variables, ya que analiza la frecuencia de los hallazgos más que una medición *per se*.⁴

¿En qué consiste una base de datos?

Una base de datos engloba el conjunto de información que mantienen una relación entre sí, las cuales se organizan con la finalidad de facilitar su posterior análisis.¹⁰ Entre las principales funciones y ventajas de una base de datos, cabe destacar su capacidad de

Tabla 1: Clasificación y ejemplos de los tipos de variables.¹⁻⁷

Característica	Tipo de variable		Descripción	Ejemplo
Naturaleza	Cuantitativo	Discreto	Representados en números enteros y finitos (sin decimales)	Número de embarazos (No es posible tener 1.5 embarazos)
		Continuo de intervalo	El cero no se considera como valor absoluto, por lo que abarca números tanto positivos como negativos	Temperatura (centígrados)
		Continuo de razón	El cero se considera un valor absoluto, por lo que las variables sólo adoptan números positivos	Peso (kilogramos)
	Cualitativo	Nominal	Las variables se nombran según su característica cualitativa sin que las categorías a las que pertenecen sean más importantes que otras	• Dicotómico (dos categorías): Sí/ No
		Ordinal	Las variables se presentan en orden jerárquico con intervalos graduales que pueden ser equidistantes o no	• Policotómico (> 2 categorías): Tipo de sangre (A, B, O, AB)
				Escala de dolor: leve, moderado, severo
Relación o función	Dependiente		Resultado de la intervención de la variable independiente, la cual solo es observada por el investigador	Complicaciones postquirúrgicas
	Independiente		Variables asignadas y controladas por el investigador que pueden producir cambios en otra variable	• Cantidad de sangrado
	Interviniente		Produce efecto directo sobre las variables dependientes	• Tasa de infección del sitio quirúrgico
	De confusión		Distorsiona la asociación entre variables, resultando en errores, sesgos y dudas	• Tipo de abordaje quirúrgico
				• Abierta
				• Mínimamente invasiva
				Comorbilidades
				• Diabetes mellitus
				• Hipertensión arterial sistémica
				• Consumo de anticoagulantes
				• Edad
				• Cuidado ambulatorio de la herida quirúrgica
				• Apego a las indicaciones médicas de parte del paciente

depuración de datos y disminución de la redundancia o repetición de la información, prevención de la pérdida de información, facilita el análisis de un grupo de datos, disminuye el espacio de almacenamiento, agiliza y automatiza la búsqueda de información y confiere mayor flexibilidad para introducir o actualizar el contenido.^{2,11} Para construir un adecuado diseño de la base de datos se recomienda seguir las siguientes indicaciones:^{2,12}

- Definir una lista de los datos que van a ser requeridos para el estudio. Para realizar una adecuada descripción de la población que se va a estudiar, es importante incluir datos demográficos como edad, sexo, antecedentes de importancia, perfiles laboratoriales, entre otros.
- Identificar las claves principales para evitar la redundancia de datos. Esta sección puede abarcar el número de expediente o de identificación del paciente, ya que estos datos son exclusivos de cada

individuo y así es posible prevenir la duplicación de la información.

- Personas y colaboradores recomiendan construir una base de datos en plataformas digitales para permitir al mismo paciente introducir la información, así reducir la carga de trabajo en el personal de salud y reducir el riesgo de registrar datos incorrectos al momento del interrogatorio.

Operacionalización de las variables

La operacionalización de las variables es el proceso mediante el cual éstas se vuelven “medibles” mediante un proceso de clasificación y categorización de los datos, dependiendo de sus características.¹³ Una vez establecidas las variables que se desea analizar, es necesario definir los instrumentos que se emplearán para medirlos, además de la descripción de los procedimientos e intervenciones que sufren dichas variables

a lo largo del trabajo de investigación.¹⁴ Existe una infinidad de maneras de operacionalizar una variable y la calidad en la cual se realiza dependerá completamente de los alcances del investigador. Una vez definida la naturaleza de las variables, es importante estandarizar los datos que se pretende ingresar al sistema,¹⁰ esto con la finalidad de homogeneizar la información, reducir errores al ingresar datos y permitir un adecuado conteo de los hallazgos. Si una variable no está bien “operacionalizada”, su calidad de medición también se volverá más deficiente y como consecuencia, los datos obtenidos con base de dicha variable también carecerán de calidad y el trabajo de investigación perdería su valor.¹⁵ En la *Tabla 2* se especifica un ejemplo de operacionalización de las variables, junto con sus componentes y aplicación en un estudio científico.

Cómo estructurar una base de datos

Una herramienta sencilla y básica que funciona como base de datos es la hoja de cálculos frecuentemente

utilizada para una preparación preliminar de los datos. La aplicación más reconocida para este trabajo es Microsoft Excel. En la hoja de cálculo, cada fila corresponde a un paciente y cada columna representa una variable. En ocasiones es conveniente codificar las variables cualitativas con la finalidad de facilitar su manejo, por ejemplo, asignar los valores numéricos dependiendo del origen del paciente cuyos datos se están registrando: 1 (consulta externa), 2 (urgencias), 3 (hospitalizado), entre otros. En el caso de las variables cuantitativas por lo general no requieren ser asignadas a códigos, a excepción de los casos donde las variables se organizan en rangos, por ejemplo, el tiempo quirúrgico: 1 (1 a 2 horas), 2 (3 a 4 horas), 3 (> 5 horas) y así sucesivamente.^{2,9} En la *Figura 1* se describe un breve ejemplo de la estructura de una base de datos plasmada en una hoja de cálculo.

CONCLUSIÓN

La vasta cantidad de información que debe abarcar un trabajo de investigación requiere de las herramientas

Tabla 2: Aplicación de la operacionalización de las variables.

Variable	Definición	Tipo de variable	Escala de medición	Ánálisis estadístico
Número de identificación	Número de afiliación del paciente que pertenece a la institución	Identificación	NA	NA
Edad	Años cumplidos desde fecha de nacimiento	Cuantitativo discreto	Años	Promedio Desviación estándar
Tiempo quirúrgico	Tiempo total abarcado por el procedimiento quirúrgico	Cuantitativo continuo	Horas	Promedio Desviación estándar
Dolor postquirúrgico	Intensidad del dolor referido por el paciente después del procedimiento quirúrgico	Cualitativo ordinal	Leve, moderado, severo	Frecuencias Prueba de Fisher

NA = No aplica.

G1					
	A	B	C	D	E
	Número de identificación	Edad	Origen	Tiempo quirúrgico	Dolor posquirúrgico
Paciente 1	1	21	2	1	Leve
Paciente 2	2	43	3	1	Moderado
Paciente 3	3	38	1	2	Leve
Paciente 4	4	57	1	1	Leve
	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5

Figura 1: Estructura básica de una base de datos en una hoja de cálculo.

tas más adecuadas para un análisis apropiado. Las bases de datos son pieza fundamental en el proceso de investigación, ya que permiten la organización, el almacenamiento y la gestión de grandes volúmenes de información y con ello facilitan la interpretación de los datos más relevantes del trabajo. Al mejorar la precisión y reproducibilidad del trabajo de investigación, proporcionan un método estructurado y sistemático para dicho análisis. Se trata de un excelente instrumento que promueve la colaboración entre varios investigadores mediante el intercambio de datos e integración de la información de diferentes fuentes, ampliando todavía más la generación de nuevos descubrimientos y los alcances del estudio. Una base de datos con un diseño adecuado garantiza que otros investigadores puedan reproducir el trabajo de investigación y contribuir en el avance del conocimiento científico.

Una base de datos debe ser fácil de utilizar para permitir a todos los investigadores involucrados tener un acceso apropiado. Los puntos más importantes a recordar para estructurar una adecuada base de datos son tener bien definido el objetivo principal del trabajo de investigación, y por tanto, la base de datos que va a ser estructurada por las variables, las cuales deben estar bien delimitadas para ir de la mano con la pregunta de investigación. Se recomienda crear la base de datos a partir de un esquema de diseño que incluya variables estandarizadas y las directrices sobre cómo registrar la información. Una vez iniciado el proceso de recopilación de datos es importante realizar monitoreos constantes para identificar y corregir posibles errores o inconsistencias, además de mantener la información actualizada para prevenir la pérdida de los mismos. Una vez concluido el proceso de recolección y análisis, el reporte de los resultados obtenidos debe ser comprensible y accesible para todos los integrantes del equipo de investigación y otras partes interesadas.

REFERENCIAS

1. Hazra A, Gogtay N. Biostatistics series module 1: basics of biostatistics. Indian J Dermatol. 2016; 61: 10-20.
2. García J, López J, Jiménez F, Ramírez Y, Lino L, Reding A. Metodología de la investigación, bioestadística y bioinformática en ciencias médicas y de la salud. 2da edición. México: McGRAW-HILL; 2014.
3. Espinoza E. Las variables y su operacionalización en la investigación educativa. Parte I. RC. 2018; 14: 39-49.
4. Mayya S, Monteiro A, Ganapathy S. Types of biological variables. J Thorac Dis. 2017; 9: 1730-1733.
5. De Muth J. Overview of biostatistics used in clinical research. Am J Health-Syst Pharm. 2009; 66: 70-81.
6. Cash E, Boktor S. Understanding Biostatistics Interpretation. 2023. In: StatPearls. Treasure Island (FL): StatPearls Publishing.
7. Guzmán J, Gutiérrez E, Hernández G, Curiel J, Patiño C, Jiménez J, et al. Manual de análisis estadístico y uso de base de datos. 1ra edición. México: Universidad de Guadalajara; 2015.
8. Martínez M, Briones R, Cortés J. Metodología para la investigación para el área de la salud. 2da edición. McGraw-Hill Education; 2013.
9. Falavigna A, Jiménez J. Educación en Investigación: de la idea a la publicación. 2da edición. Brasil: EDUCS; 2015.
10. Torrachi E, Córdoba A, Chiriboga G, Villavicencio E. Estrategia de análisis de datos (Parte 1): Creación de bases de datos para investigaciones en ciencias de la salud. Odontol Act Rev Cient. 2019; 4: 13-20.
11. Malavassi A. Las bases de datos como herramienta para la investigación histórica. Diálogos Rev Electr Hist. 2012; 13: 193-197.
12. Persons J, Thomas C, Hsiao M, Courry R. How to build a research database from data you collect to guide your clinical work. Cogn Behav Pract. 2023; 30: 35-44.
13. Arias González JL, Universidad Católica de Santa María, Arequipa, Perú. Guía para elaborar la operacionalización de variables. Espacio I+D. 2012; 10: 42-56.
14. Mateu L, Huguet J, Errando C, Breda A, Palou J. Cómo escribir un artículo original. Actas Urol Esp. 2018; 42: 545-550.
15. Andrade C. A student's guide to the classification and operationalization of variables in the conceptualization and design of a clinical study: part 1. Indian J Psychol Med. 2021; 43: 177-179.

Conflictos de intereses: los autores no tienen ningún conflicto de intereses en relación a este manuscrito.