EL PRIMER ACERCAMIENTO A LA BIOESTADÍSTICA: CLASIFICACIÓN DE LAS VARIABLES

THE FIRST APPROACH TO BIOSTATISTICS: CLASSIFICATION OF VARIABLES

Richard Shoemaker *.

RESUMEN

La variable es determinada característica o propiedad del objeto de estudio que puede tomar diferentes valores. Las variables se pueden clasificar de diferentes maneras según la información que contengan y determinan la manera como se resumen. El presente artículo ilustra la definición de variable, y su clasificación como herramientas básicas para la investigación clínica.

Palabras clave: *Neonato, Histiocitosis de Células de Langerhans, masas mediastinales.*

ABSTRACT

The variable is a characteristic or property of the object of study. The variable can take different values. The variables can be classified in different ways according to the information contained, therefore it is the way that determining how are summarized. This article illustrates the variable ´s definition and its classification as basic tools for clinical research.

Keywords: *Keywords: Newborn, Histiocytosis, Langerhans cells, mediastinal mass.*

* MD, MPH, Especialista en Medicina Interna, Epidemiólogo, servicio de estadística Walter Reed Army Medical Center, Washington DC. U.S.A.

Correspondencia: Richard Shoemaker, Facultad de Ciencias de la Salud, Universidad del Cauca, Carrera 6 No. 14N-02, Popayán, Colombia. Correo electrónico: rickgshoes@hotmail.com

INTRODUCCIÓN

Lo que nos caracteriza y nos distingue a los humanos de las otras especies del planeta Tierra es nuestra gran variabilidad -biológica, cultural, psicológica y social-. Esta gran variabilidad es lo que hace difícil y a la vez necesaria las investigaciones en las ciencias de la salud. De esta característica nace la necesidad de una disciplina como la epidemiología que responda a este interrogante: ¿Por qué determinados grupos humanos son los más susceptibles a desarrollar una enfermedad?

Supongamos que nos interese recolectar información sobre la población de estudiantes de primer semestre universitario en todas las diferentes facultades de una universidad. Las características de esta población que nos interesan son: edad, género, peso, estatura, procedencia, y resultados de un examen que determina su ingreso. Estas características se llaman las variables de la *investigación*. Como su misma denominación, implica una característica que varia entre un individuo y otro y que se puede medir de alguna manera.

CLASIFICACION

Se puede clasificar variables en diferentes maneras y tales esquemas de clasificación son útiles para determinar el método de resumen y de análisis de los datos más adecuado que se debe aplicar. Miremos tres diferentes esquemas de clasificación que son los más útiles en la investigación epidemiológica.

Según su interrelación

Se clasifican en dependientes e independientes. La variable dependiente es el objeto de interés, que varía en respuesta a una intervención o exposición. La variable independiente es la intervención o exposición. Este esquema depende de los objetivos de la investigación en vez de una estructura matemática inherente de la variable misma. Si se va a describir una variable, que se esta investigando, en términos de otras variables, se llama una variable *dependiente*. Si se utiliza una variable para describir una variable dependiente entonces se llama una variable *independiente*. Por ejemplo, si nos interesa saber si la presión arterial cambia con la edad, la presión arterial sería *descrita* por la variable edad, o sea, nos interesa saber si los cambios de la presión arterial dependen de los cambios de la edad de las personas. La presión arterial sería la variable dependiente y la edad la variable independiente.

Por lo general la distinción entre variables dependientes e independientes es obvia y no presenta problemas. Sin embargo, una variable considerada como dependiente para evaluar un objetivo de estudio puede considerarse como independiente para lograr *otro* objetivo del estudio. Por ejemplo, en adición de la relación entre edad y presión arterial, en el mismo estudio, puede que me interese saber si la presión arterial se asocia con infartos cardiacos de los miembros de la muestra. La variable presión arterial ahora se convierte en una variable independiente y el infarto es la variable dependiente.

Según su contenido

Según su contenido se pueden clasificar en unidimensionales y multidimensionales. Las unidimensionales reflejan conceptualmente un solo atributo de un individuo o grupo como propiedad única cuantificable. Su medición se hace por una sola escala. Ejemplos: Peso, Talla. Indicadores (tasa de mortalidad). Las multidimensionales son combinaciones o resultado de la agregación matemática de varias variables unidimensionales para formar los llamados índices. Ejemplos. Apgar del neonato, Índice de calidad de vida, Índice de costo de la vida, Años de vida saludables perdidos.

Según su escala de medición

Las escalas de medición son categorías en las cuales se pueden dividir o agrupar las variables. Hay dos escalas de medición: cualitativas y cuantitativas. La escala de medición cualitativa tiene dos niveles de medición: Nominal y Ordinal. La escala cuantitativa también posee dos niveles de medición: Nivel de razón y nivel de intervalo. (Ver Figura 1).

Figura 1. Clasificación de las variables según su nivel de medición.

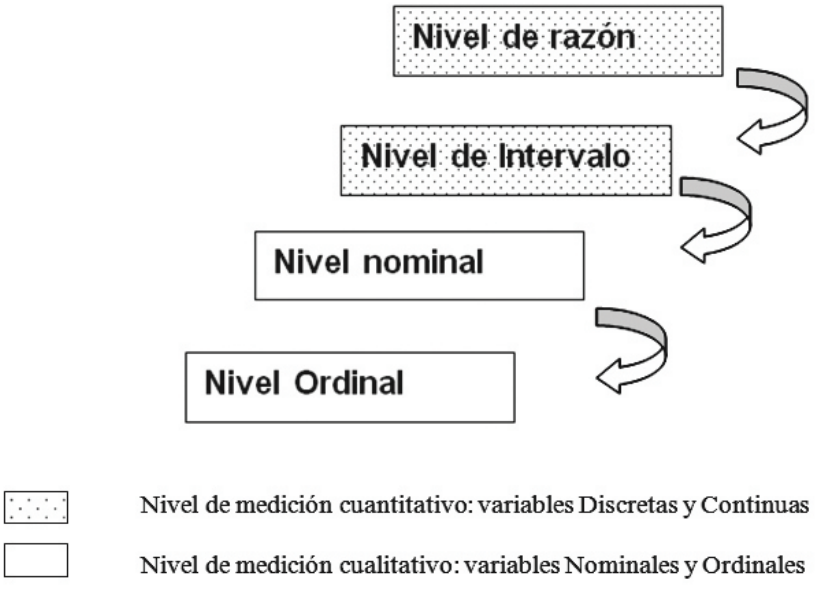
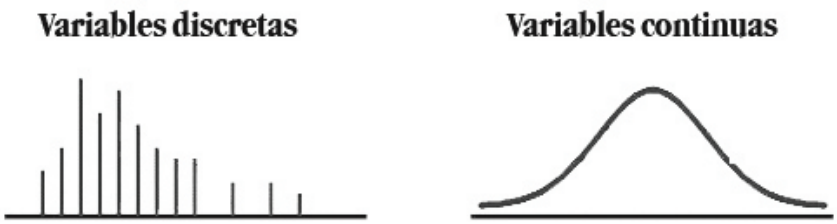


Figura 2. Clasificación de las variables cuantitativas.



Al nivel de medición cuantitativo pertenecen las variables Discretas y Continuas, al nivel cualitativo las variables Nominales y Ordinales. Pasamos a describir las características de cada una de ellas.

Variables cuantitativas: En estas las diferencias entre los elementos de la una población se expresan en cantidades o valores numéricos. Se clasifican en Discretas y Continuas. En este esquema determinamos si existen espacios entre valores sucesivos de una variable. Si existen *espacios* la variable es discreta y si no existen espacios la variable es continua. Más precisamente, es discreta la variable si, entre cualesquiera de dos valores potencialmente observables, existe un valor que no es posible observar. Una variable es continua, si, entre cualesquiera de dos valores potencialmente observables, siempre existe otro valor potencialmente observable (Ver Figura 2). Otra manera de entenderlo es que las variables discretas tienen valores de *números enteros* y las variables continuas pueden asumir cualquier valor en el intervalo entre los números enteros.

Las variables discretas toman valores que podemos suponer que son siempre números enteros. Por ejemplo el número de hijos por mujer es de 2, no puede ser 2.5. Tanto el orden como la magnitud son importantes. Se puede observar 5 nacidos vivos durante un turno de noche en el hospital pero no 5,34 nacidos vivos. Ejemplos: En clínica el mejor ejemplo son los recuentos, como el número de leucocitos por campo. Otros ejemplos: número de ingresos hospitalarios, número de hijos por mujer, número de caries dentales por escolar.

Las variables continuas pueden asumir cualquier valor en el intervalo entre los números enteros. Ejemplos: edad, presión arterial, nivel de colesterol, estatura, peso. La variable puede asumir valores sin espacios en la escala de medición, como 8.567 kilogramos, 546.7895 centímetros.

Las variables continuas tienen, a su vez, dos categorías según su medición: Nivel de Intervalo y nivel de razón o proporcionales.

Si la distancia entre valores es constante, hemos graduado a lo que se llama una variable *interválica. Una variable interválica tiene distancias iguales entre sus valores, pero el punto cero es arbitrario.* Lo que significa "punto cero arbitrario" es que el cero no tiene significación especial y, por lo tanto, podemos cambiarlo. El factor limitante del valor observado es la precisión del instrumento de medición. Para ilustrarlo, comparemos la inteligencia, medida mediante algún test de cociente de Inteligencia (CI), con alguna variable, por ejemplo el peso, donde el cero es significativo. Todos sabemos el significado de un peso cero. No podemos decidir de repente que restaremos 10kg de cualquier cosa que pesemos y decir que algo que previamente pesaba 11kg ahora pesa 1kg. Esto es algo más que una cuestión semántica; si algo pesaba 5kg antes deberíamos decir que pesa 5kg tras la conversión; una imposibilidad evidente.

Una escala de inteligencia tiene una entidad diferente. Decimos que la media del CI es 100, pero esto es solamente una convención. Pero, se puede decidir que a partir de ahora, consideramos que la media es 500, sólo con añadir 400 a todos los datos. No hemos ganado nada pero tampoco hemos perdido nada; el único cambio necesario es que ahora tendremos que ajustar de nuevo los parámetros que habíamos aprendido de acuerdo con la nueva medida.

Veamos ahora las implicaciones que ello conlleva. Como los intervalos son iguales, la diferencia entre un CI de 70 y uno de 80 es la misma que entre 120 y 130. Sin embargo, un CI de 100 no es dos veces más elevado que uno de 50. La cuestión es que si el cero es artificial (arbitrario) y móvil, las diferencias entre números son significativas, pero las razones entre ellas, no.

Si el punto cero es significativo, las razones entre números lo son también, y hemos tropezado con una variable *proporcional. Una variable proporcional o de razón tiene intervalos iguales entre valores y un punto cero significativo.* La mayoría de los tests de laboratorio son variables de razón, al igual que los atributos físicos como la estatura y el peso. Una persona que pesa 100 kg es dos veces más pesada que una que pesa 50kg; incluso si convertimos los kilos en libras, la proporción es la misma: 220 libras a 100.

Variables cualitativas: Las variables cualitativas dividen a los individuos en categorías no numéricas, mutuamente excluyentes. Se clasifican en: Nominales y Ordinales. Una variable nominal consiste en categorías a las que se les asigna un nombre sin que exista ningún orden implícito entre ellas. Pueden ser dicotómicas en el caso de tener dos alternativas, por ejemplo una variable como el género sólo puede tomar dos valores: masculino y femenino. Ninguno de los dos es más *alto o mejor* que el otro. Los podemos enumerar colocando masculino en primer lugar o femenino primero sin que se pierda ninguna información. Las categorías nominales más sencillas son las que se llaman «existenciales»: una propiedad

existe o no existe. Una persona tiene cáncer del hígado o no lo tiene, recibe o no recibe tratamiento, y la más existencial de todas, el individuo está vivo o muerto. También pueden tener un número mayor de categorías. Podemos clasificar el estado civil de una persona como soltero/ casado/ separado/ divorciado/ viudo/ unión libre (seis categorías); su color de ojos, en negro/ castaño/ azul/ verde/ mezclado (cinco categorías).

Las variables ordinales consisten en categorías ordenadas, de manera que las diferencias entre categorías pueden no ser iguales Ejemplos: Grados de insuficiencia cardiaca, grados de edema pretibial, grados de intensidad de los soplos cardiacos, estados funcionales.

Las distintas calificaciones de un estudiante dadas como excelente/ bueno/ regular/ malo tienen cuatro categorías. Difieren de una variable como la religión en el hecho de que existe una ordenación entre estos valores: «excelente» es mejor que «bueno» y éste, a su vez, mejor que «regular», que a su vez es mejor que «malo». Sin embargo, no podemos suponer que la diferencia entre «excelente» o «bueno» y «regular» sea la misma que la existente entre «regular» y «malo». Esto se ve mejor con las calificaciones numéricas; existe una pequeña diferencia ente 4,80 y 4,70; pero la diferencia es bastante grande entre 3,10 y 2,90.

Una cuestión importante, que se debe tener siempre en cuenta es que las variables cuantitativas pueden transformarse en variables cualitativas, pero las cualitativas, una vez medidas, no pueden hacerlo a cuantitativas. De tal manera que al diseñar el método de recolección de información, siempre se debe recolectar los datos de manera cuantitativa, que permite la transformación de los mismos en categorías. Por otra parte, las variables cuantitativas favorecen los análisis estadísticos, dan mayor riqueza de información y mayor poder al estudio en los estudios cuantitativos y requieren de menor tamaño muestral. Por regla general se prefiere emplear este tipo de variables porque hacen más eficientes los estudios.

Pese a que las diferencias entre datos nominales, ordinales, interválicas y proporcionales se presentan de manera natural sobre el papel, las líneas que las delimitan son a veces muy borrosas. Pese a esto, las diferencias entre variables son importantes hasta el punto de que hay que tenerlas en la mente, puesto que determinan los tipos de pruebas estadísticas que debemos aplicar a cada una de ellas.

REFERENCIAS

1. Norman, G; Streiner, D., Bioestadística. Barcelona. Mosby/Doyma Libros. 1996.
2. Hennekens, C., Buring, J., Epidemiology in Medicine. Boston. Little Brown. 1987.
3. Hulley S., Cummings S. Diseños de la Investigación clínica. Barcelona. Ed. Doyma. 1993.
4. Quintero, C., Herrera AN., Sanchez R., Nociones de bioestadística en Ardila, E., Sanchez R., Echeverry J. Estrategias de investigación en medicina clínica. Bogotá. Editorial Manual Moderno. 2001; 25-53.