

Tema 3: Variables. Medidas. Síntesis de datos estadísticos.

--**Variables.** Como ya se vio en el tema 1, las variables son características que se distinguen por la variabilidad con que se manifiestan en los diversos individuos.

--Tipos de variables.

Hay variables: cualitativas (CL) y cuantitativas (CT)

nombre	datos	expresión	variantes	ejemplo	
CUALITATIVAS O ATRIBUTOS	Catagóricos	modalidades o categorías	2 modalidades más de 2 mod.	sexo caras dado	mujer-hombre 1, 2, 3, 4, 5, 6
CUANTITATIVAS	métricos	valores	-continuas -discretas	talla nº hijos	170 cm. 0, 1, 2, 3,

--Medida de las variables

Se hace según las llamadas escalas. Básicamente hay 4 escalas de medidas:

- nominales
- ordinales
- de intervalo
- de razón

Las variables ordinales son una variante de las nominales y las de razón de las de intervalo.

--Escalas nominales

Se utilizan para medir atributos, es decir, variables cualitativas. Se da un nombre a cada una de las modalidades, se asignan los individuos a ellas y se cuentan los individuos de cada modalidad (frecuencia). El orden en que se designan las modalidades es indiferente, p.e. alto y bajo o bajo y alto.

Ejemplo: la variable sexo tiene dos modalidades, hombre y mujer. Medimos este atributo en 100 personas y encontramos 52 hombres y 48 mujeres.

En vez de dar un nombre convencional a las modalidades se las puede designar con un número, lo que facilita sobre todo el tratamiento informático. Estos números son realmente un nombre y por tanto no pueden hacerse con ellos operaciones matemáticas. Así podríamos llamar a los hombres "1" y a las mujeres "2" (ó 7 y 8...)

--Escalas ordinales

Una escala ordinal es una escala nominal en la que las diversa modalidades guardan entre sí una relación de orden o jerarquía, que debe ser respetada, siendo indiferente que el orden sea de mayor a menor o viceversa. Ese orden viene marcado por el sentido común y también por la costumbre.

Un ejemplo clásico son las notas académicas tradicionales : sobresaliente-notable-aprobado-suspenseo o suspenseo-aprobado-notable-sobresaliente. En la variable "evolución de la enfermedad" podríamos distinguir las siguientes modalidades : muerto-peor-igual-mejor-curado , o bien, curado-mejor-igual-peor-muerto.

También pueden emplearse números como nombre de modalidades, pero respetando el orden. Podríamos hacer muerto=1, peor=2, igual=3, mejor=4, curado=5 . O bien, curado=1, mejor=2, igual=3, peor=4 , muerto=5 .

--Escalas de intervalo

Se utilizan para medir variables cuantitativas cuando no hay cero absoluto en la zona de medición, lo que permite valores negativos. El cero se asigna arbitrariamente así como la unidad de medida.. La escala ha sido diseñada de tal manera que sus números permiten valorar exactamente la diferencia que hay entre dos medidas (= intervalo). Ejemplo típico es la temperatura medida de la forma habitual, lo que puede hacerse de diversas maneras. En Europa se mide en grados

centígrados o Celsius (C). El “0” se asigna a la temperatura de congelación del agua destilada y el “100” a su temperatura de ebullición. Ese intervalo se divide en 100 partes y así se obtienen los grados centígrados. En USA se mide en grados Fahrenheit (F). 0° C equivalen a 32° F y 0° F equivalen a -17,78° C. Por tanto 32° C no representa el doble de calor que 16° C, simplemente el doble de grados C. Esas temperaturas medidas en grados Fahrenheit serían 0° F y -8,9° F. Un niño con un proceso febril en Castellón puede tener 40° C de fiebre; en USA tendría 104° F. Por la Física sabemos que hay un mínimo infranqueable de temperatura, el llamado “cero absoluto”, que en grados centígrados corresponde a -273,15°. Este cero no significa la ausencia de temperatura, sino el mínimo de temperatura posible. La escala de Kelvin asigna su 0 a esta temperatura.

--Escala de razón

Se utilizan para medir variables cuantitativas cuando hay un cero absoluto, siendo la unidad de medida lo único arbitrario. Una longitud puede ser medida en cm., Km., yardas, varas, etc. pero el cero es el mismo para todos. El tiempo de reacción a un estímulo siempre empieza en cero cualquiera que sea el sistema que utilicemos para medir el tiempo. Aquí sí puede decirse que una persona que pesa 50 Kg. pesa el doble que un niño que pesa 25. Y que la diferencia de peso entre una persona que pese 80 Kg. y otra que pese 50 Kg. es la misma que la existente entre dos piedras de 35 y 5 Kg., respectivamente. No hay valores negativos.

--Variables cualitativas

Las variables cualitativas (CL) o atributos se miden por escalas nominales u ordinales según corresponda. Cuando sólo tienen dos modalidades se llaman dicotómicas. Ejemplos: cara-cruz, varón-hembra, vivo-muerto. Todos los atributos, con independencia del número de modalidades que tengan, pueden ser siempre reducidos a dicotómicos si así se desea. Los 4 palos de la baraja española (oros, copas, espadas y bastos) pueden ser reducidos a oros-no oros, bastos-no bastos, etc. ; las marcas de coches a Seat-no Seat. ; el estado civil a casado-no casado...

--Variables cuantitativas

Las variables cuantitativas (CT) se miden por escalas de intervalo o de razón, según su naturaleza. Pueden ser continuas o discretas.

Una variable CT es continua cuando puede tomar cualquier valor en su zona de variabilidad. Son continuas la talla, el peso, la tensión arterial, el contenido de un frasco, la glucemia, etc.

Las variables CT discretas no pueden adoptar cualquier valor, sino solamente ciertos valores.

Una familia puede tener 0, 1, 2, 3, ... hijos, pero no 3,1416 hijos. El nº de pacientes que ingresa en un hospital, el nº de ataques que sufre un paciente en un mes, el nº de cápsulas de un envase medicamentoso... son discretas.

Una variable CT continua se mide a menudo, porque resulta más práctico, de forma “discretizada”. La edad suele expresarse en años enteros, o en meses en los niños pequeños, pero no por eso deja de ser continua.

--Transformación de variables

Las variables cuantitativas pueden ser transformadas en cualitativas, dicotómicas o no, con una pérdida en la calidad de la medida, que a veces se asume si mejora la información. La talla podemos medirla en alta-normal-baja. Los valores de colesterol en mayor de 200 mg/dl - igual o menor de 200 mg/dl. Como la variable CT proporciona más información que la CL debe ser usada siempre que no sea más conveniente hacerlo de forma cualitativa.

Las variables CL en cambio no pueden ser transformadas en CT.

Las variables CL son por su propia naturaleza discretas.
Por las limitaciones de los instrumentos de medida la mayoría de las CT continuas son discretizadas.

Dos ejemplos:

---variable “INGESTION DE ALCOHOL” .

He seleccionado 4 formas distintas en orden creciente de información:

- | | |
|--|--|
| 1) abstemio – bebedor | Variable CL con dos modalidades, nominal. |
| 2) abstemio – bebedor – alcohólico | Variable CL con tres modalidades, ordinal. |
| 3) nº de copas o vasos bebidos en una semana | Variable CT discreta |
| 4) gramos de alcohol tomados en una semana | Variable CT continua |

---“ESTUDIO DE 3 TRATAMIENTOS DE LA ISQUEMIA CORONARIA”.

Considerando las variables:

- | | |
|---------------------------------|-------------------------------|
| - sexo : hombre – mujer | CL con 2 modalidades, nominal |
| - medicamento: A – B – C | CL con 3 modalidades, nominal |
| - nº ataques del día anterior | CT discreta |
| - distancia caminada sin disnea | CT continua |

--Necesidad de una definición clara de las variables

Es esencial que todo el mundo sepa qué se está midiendo y cómo. Está claro lo que es medir el peso en Kg. o la talla en cm. Pero, ¿que es ser “fumador”? ¿El que fuma un pitillo, aunque sea una vez al año? ¿O el que fuma cada día o al menos cada tres?... Hay que concretar y decir por ejemplo: “en este estudio se considera fumador a quien fuma al menos un cigarrillo cada semana” o “se considera desnutridos a los niños que en los gráficos peso/talla de Tanner están por debajo del percentil 3”, etc., etc.

--Dominio de una variable

Es el conjunto de valores o modalidades que puede adoptar. El dominio de la variable CL “puntuación de la cara de un dado” es (1, 2, 3, 4, 5 y 6). El de la variable sexo: (hombre, mujer). El de la “longitud de las hojas de la planta P” cualquier valor entre 1 y 8 cm. o $\in (1 \div 8)$, etc.

--Variables aleatorias y controladas

Una variable es controlada o independiente cuando su valor o la modalidad elegida en cada uno de los individuos depende únicamente del investigador. En un estudio podemos seleccionar sólo individuos del sexo masculino. O fijar la dosis de medicamento que se da a los ratoncillos, etc. Una variable es aleatoria o dependiente cuando su valor en cada uno de los individuos no depende del investigador, sino de la naturaleza o reacción del propio individuo. Por ejemplo la talla de los alumnos de una clase, la tensión arterial de un grupo de pacientes, etc.

--Medida de una variable continua

Debido a la imperfección de los instrumentos de medida, aún de los más sofisticados, el valor exacto o real de una medida (X_e) es realmente desconocido y sólo podemos expresarlo de una forma aproximada mediante el valor medido (X). Supongamos que estamos midiendo una longitud con una regla graduada. Cuando la medida no se corresponde con un valor marcado en la regla, hay que aproximar (=redondear) a la marca más cercana. Si hay equidistancia se aproxima al valor par.

	5	6	7	8
	-x - - - - -x - - -x - -			
medida:	5	7	8	

La diferencia entre el valor exacto y el valor medido se llama ERROR ABSOLUTO. Toda medida tiene su error.

$$\boxed{E = |X_e - X| \text{ y por tanto } X_e = X \pm E}$$

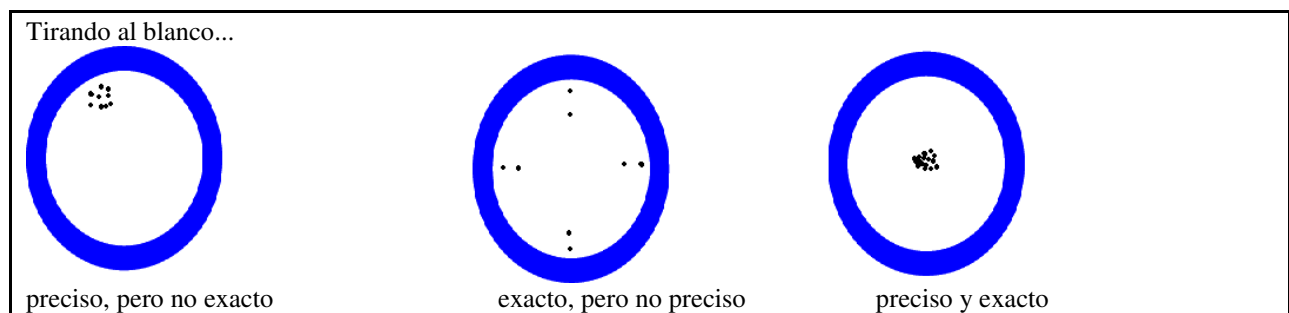
Este error, del que sólo podemos conocer su máximo (**E_{max}**), depende de la sensibilidad, precisión y exactitud de los instrumentos de medida.

La sensibilidad (*se*) es la unidad más pequeña que permite utilizar el instrumento de medida. En las reglas graduadas habituales $se = 1 \text{ mm}$. El **E_{max}** es igual a la mitad de la sensibilidad; **E_{max}** = $se/2$. Una regla milimetrada: tiene un **E_{max}** de $1/2 \text{ mm} = 0,5 \text{ mm}$.

Hay precisión cuando repetida la medida muchas veces da valores iguales o muy parecidos.

Hay exactitud si la media de repetidas medidas coincide con el valor exacto de la medida.

Así, si una longitud real de 9,0 cm. la medimos 4 veces y obtenemos 9,1; 9,0; 9,0 y 8,9 el instrumento es preciso y exacto. Si obtenemos 5,6 ; 5,5; 5,7; 5,6 será preciso, pero no exacto. Midiendo 9 ; 6 ; 12 ; 3 y 15 será exacto pero no preciso. La medida ideal es la que se obtiene con un máximo de sensibilidad, precisión y exactitud.



--¿Que sensibilidad se debe utilizar?

Una sensibilidad escasa proporciona datos de poca confianza, con mayor margen de error. Si es excesiva no es mala en sí, pero en general supone aparatos más caros y de manejo más difícil.

Hay que elegir la más adecuada para cada caso concreto, teniendo en cuenta la experiencia y el sentido común.

La sensibilidad es adecuada si la diferencia entre la medida más alta ,sin punto o coma decimal, y la medida más baja , también sin punto o coma decimal, está entre 30 y 300 .

Ejemplo:

1- medimos en mm. la longitud de las hojas de la planta XYZ. La medida mayor es 8 y la menor 4. Como $8-4=4$, que es menor de 30, la sensibilidad utilizada no es buena. En una medida de 5 mm. el error máximo es de 0,5 mm., o sea de un 10%. El instrumento de medida no es adecuado.

2- después utilizamos un aparato que mide en décimas de mm. Como valores extremos obtenemos 8,4 y 4,3 mm. $84-43=41$, que está entre 30 y 300. En una medida de 5,0 mm. el error máximo es de 0,05 mm., un 1%. Este instrumento sí es adecuado.

--Valor puntual y por intervalo de una medida

Al desconocer el valor exacto de una medida, **X_e** , hay que estimarlo. La medida se puede expresar de dos formas: puntual o por intervalo.

La medida puntual o valor puntual es el valor medido, **X** ; por tanto no es exacto..

El valor por intervalo o medida por intervalo es el intervalo en el que con seguridad (¡si se ha medido bien!) estará el valor exacto **X_e** de la medida. Se obtiene sumando y restando al valor puntual el error máximo, es decir, la mitad de la sensibilidad: $X \pm se/2$. De esta forma se obtienen los llamados *límites reales de la medida*, uno superior y otro inferior. Si medimos un lápiz con una regla milimetrada y obtenemos 151 mm., la medida puntual será 151 mm. Como la sensibilidad es de 1 mm., la medida por intervalo será $151 \pm 0,5 \text{ mm}$. o $\in (150,5 \div 151,5)$.

Si utilizamos una regla con nonius, que mide en décimas de mm. y obtenemos 151,1 mm. , el valor puntual será 151,1mm. Aquí la sensibilidad es de 0,1 y por tanto la medida por intervalo será $151,1 \pm 0,05$ ó $\in (151,05 \div 151,15)$.

Como es fácil equivocarse al realizar los cálculos, puede resultar útil el procedimiento siguiente:

- a) se toma el número, prescindiendo del posible punto o coma decimal y se añade un 0
- b) se le suma y resta 5
- c) si había decimales, se vuelve a poner la coma o punto decimal en su sitio. Así tenemos los dos límites del intervalo.

En el último ejemplo: $151,1 \rightarrow 15110 \rightarrow -5 = 15105$ y $+5 = 15115 \rightarrow 151,05$ y $151,15$

--Cifras significativas

Son las cifras del valor puntual de una medida, prescindiendo de los ceros a la izquierda de la primera cifra con valor distinto de cero. Son pues función de la sensibilidad.

medida	cifras significativas	medida	cifras significativas
65,5 m	3	4,53400 cm	6
0,0018 kg	2	1,00180 amp	6
1,0018 mm	5	0,10000 sec	5

En un número redondeado las cifras significativas llegan tan sólo hasta el lugar del redondeo. 18 millones como redondeo de 18 234 156 tiene 2 cifras significativas ; 3,14 como redondeo de 3,141592 tiene 3.

--Métodos de recuento (variables CL)

- a) observación, utilizando los órganos de los sentidos.
- b) gráficos: métodos de palotes, cuadrados...
- c) tarjetas de formas, contenidos o colores distintos
- d) lectura óptica, como en el escrutinio de quinielas y similares
- e) lectura magnética (de espacios marcados con lápiz de grafito)

--Síntesis de datos estadísticos

Una vez medida la variable en los diversos individuos se tiene una serie de datos, métricos o categóricos, los llamados DATOS ORIGINALES o DATOS AISLADOS, que sin más elaboración suelen ser poco útiles.

Es necesario ordenarlos y resumirlos para que proporcionen la máxima información de la forma más sencilla posible. Esto se hace de diversas formas:

- agrupando los datos según su frecuencia, con lo que se transforman en DATOS AGRUPADOS O DISTRIBUCION DE FRECUENCIAS, construyendo las correspondientes TABLAS y GRAFICOS ESTADISTICOS
- calculando los llamados INDICES o PARAMETROS ESTADISTICOS, como media aritmética, desviación estándar, porcentajes, etc.

Las Escuelas clásicas utilizan el término INDICE para las muestras y sus símbolos se representan con letras latinas, mientras que el término PARAMETRO se reserva para las poblaciones, con símbolos de letras griegas. Aquí utilizaremos ambos términos de forma indistinta, es decir, tanto para poblaciones como para muestras. Y salvo alguna rara excepción los símbolos serán de letras latinas.

Recordatorio : MEDIDA DE UNA VARIABLE CONTINUA

X_e	valor exacto, real, de la medida ; es desconocido
X	valor medido por el instrumento; es el valor puntual
E = X_e - X 	error de la medida ; por tanto $X_e = X \pm E$
E Máximo (E_{max})	$se/2$
Valor por intervalo de una medida	$X \pm E_{max}$ ó $\in (X-E_{max} \div X+E_{max})$ en ese intervalo está contenido el valor real X _e