

Nociones preliminares

I. Conceptos Básicos de estadística descriptiva y estimación. Curso de técnicas estadísticas aplicadas al tratamiento de datos empíricos

Muñoz Varcárcel, F.; Ramírez Díaz, L.; Torres Martínez, A.; Montes del Olmo, C.; Martín de Agar Valverde, P.

Curso 1981–82. Universidad de Murcia (materiales no publicados)

Contenidos

1 Introducción **1**

2 Tipos de variables y datos **2**

1 Introducción

1. Un individuo (objeto o caso) se caracteriza por las siguientes propiedades:

- es irrepetible
- es observable y, por tanto, objeto de estudio
- es semejante (similar, equivalente) a otros respecto a criterios fijados *a priori*;
- se comporta de acuerdo con leyes objetivas (es decir, independientes del observador)

2. Una *población* (biológica, física, geográfica, sociológica, etc.) es un conjunto (colección, colectivo, repertorio, etc.) de individuos con propiedades comunes, delimitados en el espacio y en el tiempo.

Los términos individuo y población dependen en última instancia del observador. Así, la ciudad de Murcia es una población que agrupa a todos los murcianos (si estamos interesados en estudiar las características de estos), y es al mismo tiempo un individuo de la población formada por todas las capitales de provincia de España.

3. Un *experimento* es cualquier proceso (estrategia, mecanismo) que permite asociar a cada individuo de una población un símbolo, o grupo de ellos, numérico o no, de entre todos los símbolos de un conjunto dado *a priori* (+/—, colores, letras, números, etc.). Son experimentos, p.e.: la determinación de la edad y el género de los habitantes de una ciudad, la obtención de la velocidad o el rendimiento de una reacción química, la realización de protocolos de enfermos, la medición de factores ambientales, etc.

La realización concreta de un experimento recibe el nombre de *ensayo* o *prueba*, y el símbolo (o los símbolos), obtenido del ensayo o prueba se llama *resultado*, *observación* o *dato*.

4. Un *experimento aleatorio* o de azar es aquel que verifica las siguientes propiedades (Viedma, 1976):

- (a) Se puede repetir un gran número de veces bajo un conjunto de condiciones estables (repetibilidad indefinida).
- (b) En cada prueba no se puede predecir el resultado que se va a obtener (condición de azar).
- (c) La frecuencia relativa de cada resultado, es decir, el cociente de dividir el número de veces que ocurre el resultado entre el número total de pruebas, tiende a estabilizarse en torno a un número fijo, cuando el número de pruebas crece indefinidamente, comprendido entre cero y uno, que recibe el nombre de *probabilidad* de dicho resultado (ley de regularidad estadística).

Son experimentos aleatorios, p.e.: la determinación del número de huevos depositados por una hembra de *Drosophila melanogaster*, recuento de las defunciones que se darán en una ciudad en un año determinado, la observación del porcentaje de enfermos que sanarán tras aplicarles una u otra droga, etc.

5. Un experimento determinista es aquel en el que los resultados están totalmente determinados una vez que se fijan las condiciones de realización del mismo (condiciones iniciales). En un experimento determinista, se mantiene la condición (a), no se mantiene la (b), y la frecuencia relativa mencionada en (c) se mantiene prácticamente igual a la unidad.

Son experimentos deterministas: el cálculo de la velocidad de un objeto en caída libre después de transcurrido un tiempo t , la determinación de la posición de la luna en una fecha futura, el peso de agua obtenida de cantidades conocidas de hidrógeno y oxígeno, etc.

En sentido estricto, ningún experimento es determinista, ya que, siempre existen discrepancias entre los resultados previstos por la teoría y los obtenidos en la práctica. En general, estas discrepancias suelen atribuirse a algunas de estas causas: insuficiencia del modelo teórico, errores de observación/medición, defectos en los aparatos de medida, etc.

6. El conjunto (colección, colectivo) de símbolos asociables a cada individuo de una población por un experimento se llama *espacio muestral*, *población estadística*, o *suceso seguro*, y se representa por la letra ω . Por ejemplo, en la determinación del sexo en los humanos el espacio muestral está reducido al conjunto de macho/hembra o, en el lanzamiento de un dado, el espacio muestral está limitado por el conjunto de símbolos 1, 2, 3, 4, 5 y 6.

Cada elemento de ω recibe cualquiera de estos nombres: *observación*, *resultado* o *dato*.

7. Una *variable* es cualquier característica que podemos observar en un individuo, p.e.: peso, temperatura, número de delitos, etc.
8. Una *muestra* es una parte del espacio muestral. Es, por tanto, una colección de observaciones o datos. Cualquier proceso de obtención de muestras recibe el nombre de muestreo. El número de elementos que forman la muestra se llama *tamaño muestral* o *de la muestra* (n).

2 Tipos de variables y datos

En el esquema 1 se recogen las relaciones entre los distintos tipos de variables, que a continuación se desarrollan.

1. *Variables cuantitativas* son aquellas que toman valores numéricos: pesos, longitudes, densidades, proporciones, número de nacimientos o de habitantes, etc.

Cuando la variable toma valores numéricos fijos sin valores intermedios se llama *discreta*. Cuando el espacio muestral de una variable discreta contiene una cantidad finita de elementos (por ejemplo, 1/2, 1, 3/2, 2, 7) la variable se llama *acotada*. En caso contrario es *no acotada*.

Las variables discretas se caracterizan por los saltos o interrupciones en su campo de variación. Así, el número de hijos de una pareja es una variable discreta que puede tomar los valores 0, 1, 2, 3, 4, ..., pero ningún valor entre 0 y 1, entre 1 y 2, etc. Son también discretas las variables: número de hojas de una flor, número de habitantes de una ciudad, etc.

El número de lanzamientos de un dado que hay que realizar antes de obtener por primera vez un 6 es un ejemplo de una variable *discreta no acotada*.

Las variables cuyo espacio muestral es todo un intervalo (a, b) se denominan *continuas*. Se caracterizan por tomar "todos" los valores comprendidos entre a y b , es decir, si x es cualquier valor mayor que a y menor que b hay, por lo menos, un individuo de la población para el que la variable toma el valor x . Así pues, una variable continua no presenta "huecos" en su campo de variación.

Por extensión, se llaman también continuas aquellas variables cuyo espacio muestral es la unión de varios intervalos que no se superponen.

Los pesos y las longitudes son variables continuas (si se supone que no existe límite de precisión en el aparato de medida).

2. *Variables cualitativas* son aquellas cuyo espacio muestral es un conjunto de cualidades o atributos (colores, actividad profesional, presencia o ausencia de una planta en una parcela, estados de agresividad, etc.).

Cuando es posible ordenar los valores de una variable cualitativa con algún criterio (raro, frecuente, abundante, por ejemplo), la variable se llama *ordinal*. En caso contrario, la variable es *nominal* (nombres de especies, clase de enfermedad).

La escala elegida y el grado de precisión en las observaciones influyen en la calificación de las variables. Así, si el aparato de medida utilizado para medir longitudes de personas sólo aprecia centímetros, la variable se considerará, a efectos prácticos, discreta con valores 0, 1, 2, ..., 250 centímetros (aunque “potencialmente” sea continua).

De la misma forma, una variable cualitativa “color” puede considerarse cuantitativa continua si cada clase de color se caracteriza por su longitud de onda.

3. Cuando de cada individuo de una población se observan dos o más características, el experimento correspondiente es multivariante. Cada realización del experimento conduce a un dato múltiple o conjunto de datos simples. Las variables asociadas a este tipo de experimentos se llaman *multivariantes*:

Si el experimento consiste en la determinación del peso, de la altura y del color de ojos de cada individuo de una población, cada realización dará lugar a una terna ordenada (X, Y, Z) , en la que X es el peso del individuo observado e Y y Z serán, respectivamente, su altura y su color de ojos.

Las variables univariantes X , Y y Z son las componentes de la variable trivariante (X, Y, Z) .

Con más precisión: una *variable multivariante* es un conjunto ordenado de variables univariantes que se llaman componentes y se representa por (X_1, X_2, \dots, X_k) . Cada componente puede pertenecer a cualquiera de los tipos definidos en los apartados anteriores.