

Estadística descriptiva e inferencial

En general, las técnicas estadísticas se suelen dividir en dos categorías formando las dos ramas tradicionales y básicas de la Estadística:

- Técnicas descriptivas -> Estadística descriptiva
- Técnicas inferenciales -> Estadística inferencial

La estadística descriptiva o deductiva trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones. Por lo general, cuando hay un gran número de datos antes de que la mente humana pueda interpretarlos, deben resumirse o reducirse. La mente humana no tiene la capacidad de extraer conclusiones de una serie de datos en bruto; la estadística descriptiva sirve de herramienta o instrumento para describir, resumir o reducir las propiedades de un conjunto de datos para que se puedan manejar. En definitiva, comprende aquellas técnicas que se usan para resumir la información (largas listas de valores) para proporcionar índices simples y comprensibles y, por lo tanto, para facilitar descripciones y comparaciones, haciéndolo de la forma más exacta posible. Para reducir la información se construyen tablas, se representan gráficos y se calculan parámetros estadísticos que caracterizan la distribución, de esta forma se simplifica la complejidad de todos los datos que intervienen en la distribución.

Por lo tanto, la estadística descriptiva hace referencia, o se utiliza en las etapas 3, 4 y 5 del método científico (observación, clasificación y descripción), y únicamente se limita a realizar deducciones directamente a partir de los datos y parámetros obtenidos.

Por su parte, la estadística inferencial o inductiva trata de llegar a conclusiones que sobrepasan el alcance de los datos analizados, es decir, se trata de técnicas que se emplean para inferir o deducir características desconocidas a partir de un conjunto de datos conocidos, apoyándose fundamentalmente en el cálculo de probabilidades. Esto es así porque es imposible en la mayoría de los casos utilizar como datos para comprobar las hipótesis la totalidad de los fenómenos que componen la población objeto de estudio. (Por ejemplo, es imposible medir el diámetro de todos los cantos en una terraza fluvial, o encuestar a todos los agricultores de una determinada zonas...). Como resulta imposible examinar la población entera de los fenómenos que estudiamos, la construcción de leyes y teorías se tiene que apoyar en DATOS MUESTRALES. A partir de unos pocos datos conocidos (los de la muestra), se trata de obtener información de la población total, y esto lo hace apoyándose en el cálculo de probabilidades, como hemos mencionado anteriormente.

Uno de los principales objetivos de la estadística inferencial es estimar las propiedades de una población a partir del conocimiento de sólo una muestra de ella. La estadística inferencial se basa por lo tanto en la estadística descriptiva, ya que la inferencia o deducción de las propiedades de la población entera se deriva de las características de la muestra que es analizada con las técnicas de la estadística descriptiva. En realidad su campo de acción es más amplio, tal como veréis en la segundo cuatrimestre cuando os dediquéis por completo a esta parte de la estadística.

Antes de terminar este tema introductorio es necesario dejar algunos conceptos claros.

5- Algunos conceptos básicos

Estadística: (Spiegel) estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis.

En estadística se denomina Población al conjunto de seres u objetos acerca de los que se desea tener información, (No pensar siempre en personas: el ejemplo de todos los cantos de una terraza que veíamos antes, o también la ganadería ovina de la P.I., o la población votante de un pueblo, o el parcelario agrario de un municipio si quiero estudiar su estructura...).

Se conoce como Individuo o elemento de una población a cada uno de los miembros de esa población. Los elementos de la población se llaman individuos debido al origen demográfico de esta ciencia, también lo podéis encontrar referidos como unidades estadísticas.

Muestra: subconjunto de esa población (al que sometemos a un verdadero análisis). El número de elementos de una muestra se denomina tamaño.(ej. cantos de una terraza).

Se denomina Variable estadística o simplemente variable a cada una de las características o atributos que se miden en los individuos de una población. Ej, años de edad, talla, diámetro de los cantos, estado civil, el peso de la ganadería ovina.... Las variables estadísticas presentan modalidades o valores, que son las diferentes manifestaciones de una variable. Por ejemplo, el sexo es una variable que tiene dos modalidades (también llamadas categorías): masculino y femenino. El número de hijos de un matrimonio es una variable que puede presentar los valores 0, 1, 2, 3...

Los símbolos empleados para representar una variable suelen ser las tres últimas letras del alfabeto x,y,z. El símbolo x_j ó x_i denota cualquiera de los n valores que puede tomar una variable; x_1 representa el primer valor observado, x_2 el segundo... x_i .

Notaciones: Aclarar que existen las diferentes notaciones para significar lo mismo (ej. Frecuencia: f ó h ó n ...)

Se distinguen **dos tipos principales de variables** (según el valor que puedan adoptar): cuantitativas y cualitativas

a) Cuantitativas

Miden alguna cualidad o atributo "cuantificable" de los individuos. (=>Valor numérico)

Esa cualidad puede ser *numerable*, como el número de hijos de una familia, con lo que tendríamos entonces una variable *cuantitativa discreta* o *medible*, como la superficie de una parcela, dando lugar a una variable *cuantitativa continua*.

Las *variables discretas*, al ser numerables, pueden tomar una serie de valores determinados, pero no los valores intermedios. Por ejemplo, el número de cabezas de ganado de una explotación puede ser 50 ó 51 pero no 50,5.

Las *variables continuas*, las que hemos definido como aquellas medibles, pueden tomar infinitos valores dentro de un campo de variación. Al menos en teoría puede tomar cualquier valor entre dos valores determinados, por ejemplo una persona que pese entre 65 y 66 kilos podemos suponer puede pesar cualquier número de kilos comprendidos

entre 65 y 66 (p.e. 65,325...). En realidad las variables continuas son una idealización; cualquiera que sea el procedimiento de medida que se emplee, tendrá un límite de precisión. Ese límite marca la imposibilidad de dar una medida continua y supone que en la práctica, todas las medidas son discretas. Sin embargo para los razonamientos teóricos, conviene considerar ciertas medidas como continuas.

b) Cualitativas

Se refieren a los aspectos no medibles o cualidades de los individuos (ej. sexo, estado civil, ...). Las variables cualitativas toman valores que no pueden ser ordenados ni operados conforme a las reglas aritméticas; a estos valores se les suele denominar modalidades. En el caso del estado civil ¿cuáles serían las modalidades?: soltero/a; casado/a; viudo/a; separado/a...

("valores" para las cuantitativas y "modalidades" o "categorías" para las cualitativas)

Las variables cualitativas pueden también clasificarse en 2 grupos según su "nivel de medición" (nivel de información que ofrecen):

a) Nominal

Es el nivel de medición más simple y permite la clasificación de los individuos u objetos en clases o categorías meramente descriptivas, que deben ser exhaustivas y mutuamente excluyentes. Ej. lugar de nacimiento, sexo, estado civil, color de coche...

En la escala nominal las diferentes categorías determinadas sirven simplemente de nombre o identificador (por ejemplo, las regiones de España). Si se dan valores numéricos a las diferentes categorías, el número será sólo un identificador, carente de cualquier otro significado (por ejemplo el código de un municipio, o el del DNI de una persona). A los datos proporcionados por la escala de medida nominal sólo se les puede aplicar estadísticas simples: número de casos en las diferentes categorías, proporción de casos en cada clase, etc.

b) Ordinal

Es el nivel inmediatamente superior al anterior, pues además de clasificar, ordena según se posea en mayor o menor grado la característica que se pretenda medir. Por lo tanto, ordena las categorías y establece un rango. Se trata de variables cuyos valores pueden ser ordenados de mayor a menor. Supongamos, por ejemplo, que preguntamos a un conjunto de personas sus preferencias sobre cinco localidades andaluzas en las que les gustaría veranear. Estas personas podrían responder: 1 Marbella 2 Málaga 3 Sevilla 4 Cádiz 5 Granada. Este tipo de información ordenada, clasificada de más a menos en este caso, se llama ordinal. Por lo tanto con esta escala se establecen categorías (como con la nominal), pero además es posible ordenar las distintas categorías de manera que sea posible decir que una de ellas ocupa un rango superior en la escala. [Relación con los datos cualitativos]

Con los datos ordinales se puede obtener algún tipo de medidas estadísticas como la mediana, los percentiles y coeficientes de correlación.

Además, dependiendo del proceso de obtención las variables pueden ser:

- Fundamentales: si son producto de una sola observación (superficie, talla, peso, sexo...)
- Derivadas: si son resultado de dos o más observaciones (caso de los índices como la renta per cápita, la densidad...)

Observación: el conjunto de modalidades o valores de cada variable medidos en un individuo constituye una observación. (Coincide con el n° de individuos o elementos de una población. Para nosotros 1 observación= 1 variable aunque 1= varias variables medidas en un individuo). Veamos un ejemplo:

Si de cada una de las 100 personas que constituyen la población mayor de edad de un pequeño pueblo, anotamos su *edad*, su *sexo* y si *votó* o no en las últimas elecciones generales, se tendrá un conjunto de datos con tres variables: "edad", "sexo" y "voto". Si se cuenta en años la variable edad adoptará los valores: 18, 19...; la variable sexo tendrá las modalidades V y M, y la variable

voto adoptará las modalidades S/N. Así los datos correspondientes a una persona cualquiera podrían ser: (45, V, S). Es decir, un varón de 45 años que sí votó en las últimas elecciones.

Este conjunto de tres variables medidas en un sólo individuo constituye una observación. Los datos de la población se resumirían en 100 observaciones de tres variables.

Redondeo de valores: cuando nos enfrentamos con muchos datos y cálculos complejos resulta útil el redondeo de un número hasta el entero o decimal más próximo, usando los siguientes criterios; se desprecia el último dígito si es menor que 5, y si es mayor que 5 se suma una unidad a la cifra anterior.

Aproximación a un entero: 6,2= 6

Aproximación a la décima: 5,17= 5,2

Aproximación a la centésima: 3, 177= 3,18

El problema aparece al redondear los números terminados en 5, pues o bien se puede sumar uno al número anterior o bien despreciarlo. Para no acumular errores lo que se puede hacer es sumar cuando el dígito anterior es impar y no cambiarlo cuando el número anterior es par (de esta forma se irán compensando las cantidades). Esto se llama redondeo según criterio del par más próximo (o del “entero par”). Por ej.

8,775 se redondea a 8,78 y

57,65 quedaría en 57,6 por tratarse el 6 de un número par

3'1432= 3'14= 3'1

3'1456= 3'15= 3'2