

Tema 1. Introducción

- Objeto de la estadística
- Tipos de variables
 - Variables cualitativas :
 - nominales y ordinales
 - Variables cuantitativas:
 - discretas y continuas
 - Series temporales, datos transversales, datos de panel
- Población y muestra
- Etapas en el análisis estadístico

Tema 2. Distribuciones de frecuencias

- Variable estadística ($X: x_i$)
- Distribuciones de frecuencia (x_i, n_i)
 - Frecuencia absoluta (n_i)
 - Frecuencia relativa ($f_i = n_i/N$)
 - Frecuencia acumulada (N_i)
 - Frecuencia relativa acumulada (F_i)
- Variables con datos agrupados por intervalos ($Li-1:Li$)
 - Marca de clase ($x^*i = (Li+Li-1)/2$)
 - Amplitud del intervalo ($ci = Li-Li-1$)
- Representaciones gráficas
 - Variables cualitativas
 - Diagramas sectoriales
 - Variables cuantitativas discretas
 - Diagrama de barras
 - Diagrama en escalera (acumulado)
 - Variables cuantitativas continuas (datos agrupados en intervalos)
 - Histograma. Densidad de frecuencias ($di = n_i/ci$)
 - Polígono de frecuencias (acumulado)

Bibliografía

- Teoría
 - MARTÍN-GUZMAN, P; TOLEDO, I.; LÓPEZ ORTEGA, F.J. y BELLIDO, N. "Manual de estadística: descriptiva" Ed. Thompson Civitas. 2006. Capítulos 1 y 2.
 - MARTIN-GUZMAN, P; et al. "Curso básico de estadística Económica". Ed. AC. 3ª edición. 1989. Capítulos 1 y 2.
- Práctica
 - RUBIO, L.; MARCO, R. "Problemas resueltos de Estadística Descriptiva". Ed. Nueva Visión. 2004. Capítulo 1.

TEMA 1

1.1.- Definiciones de Estadística.

1.2.- Conceptos básicos.

1.2.1.- Población: Elemento, tamaño y caracteres.

1.2.2.- Variable. Tipos de variables. Escalas de medida.

1.2.3.- Tipos de datos: sección cruzada, serie de tiempo y panel de datos.

1.3.- Muestra.

1.4.- Estadística Descriptiva e Inferencia Estadística.

1.5.- Principales etapas de un estudio estadístico.

1.1.- Definiciones de Estadística.

La Estadística se configura como la tecnología del método científico que proporciona instrumentos para la toma de decisiones cuando éstas se adoptan en ambiente de incertidumbre; siempre que esta incertidumbre pueda ser media en términos de probabilidad.

Por ello, la Estadística se preocupa de los métodos de recogida y descripción de datos, así como de generar técnicas para el análisis de la información.

Como resumen podemos decir que " La Estadística es la técnica del método científico, proporciona instrumentos y técnicas para los investigadores, que pueden ser de aplicación completamente general y útiles en cualquier campo de la ciencia" (ya sean estudios demográficos, económicos, sociales, biológicos, psicológicos, químicos, técnicos, etc.).

Ej: en *Economía*, las estadísticas son de vital importancia para llegar a conocer de manera cuantitativa las macromagnitudes económicas que nos van a definir la situación económica del país, permitiendo que se puedan elaborar las cuentas nacionales básicas para poder analizar la situación económica y poder tomar las medidas pertinentes en política económica. En *Sociología*, las Estadísticas sociales nos permiten comprender y evaluar apropiadamente la realidad social. En *Psicología*, la Estadística se utiliza para el estudio del comportamiento de los individuos.....

1.2. - Conceptos básicos.

Supongamos que vamos a trabajar con una encuesta de opinión. Queremos saber ciertas cosas sobre los alumnos del mismo curso que vosotros. Os preguntaremos algunos datos e intentaremos plantear preguntas interesantes sobre vosotros mismos, que quizás podamos tomar como referente para responder a otras sobre un conjunto más amplio de personas. Por ejemplo:

¿Cuál es la altura más habitual en tu clase?

¿Se puede considerar que tu paga es normal en comparación con la de otros chicos? ¿la mayoría tiene más o menos paga que tú?

¿Cuántos hacéis deporte regularmente? ¿Y cuántos desayunan antes de venir?

¿Qué coméis más: fruta, leche, legumbres, café, carne, pescado . . . ?

1.2.1. - Población: Elemento, tamaño y caracteres.

Antes de comenzar a responder preguntas, tenemos que precisar algunas cosas. ¿Sobre quiénes queremos obtener información? Ya hemos visto que sobre los alumnos de tu curso, luego la población no sois sólo vosotros, sino todos los alumnos de tu nivel. Pero nos llevaría mucho tiempo preguntaros a todos, y hemos decidido tomar un grupo representativo de todos los grupos de tu nivel, que en este caso sois vosotros, el Grupo 114. Así, vosotros sois la *muestra*. Además, a cada elemento de la *población* lo llamaremos *individuo*. Hagamos algunas observaciones sobre lo que acabamos de decir. Lo primero es que nosotros podemos querer estudiar alguna característica en animales, plantas o cosas, por ejemplo, la duración de las baterías en los teléfonos móviles, y en este caso, la población no sería "humana", sino que serían los diferentes modelos de teléfonos móviles.

Además, podemos encontrarnos con casos en los que la utilización de muestras esté más justificada aún que en nuestro caso, por diferentes motivos: si queremos conocer lo que votarán los españoles en las próximas elecciones, no podemos preguntarle a todos los españoles mayores de 18 años, porque serían millones de personas y supone mucho dinero y tiempo. Para estudiar, por ejemplo, la durabilidad media de unas determinadas bombillas hasta que se funden, no podemos examinarlas todas, porque cada examen supone que una bombilla se funda, es decir, es un caso en el que el individuo se destruye. Por tanto, en muchas situaciones, el muestreo está justificado

por razones económicas, de tiempo o de destrucción con el estudio de los individuos de la población.

Como resumen teórico decimos que:

La Estadística no se ocupa de hechos aislados, como la renta de una familia, la estatura de un individuo, etc, sino que se ocupa de un conjunto de datos, lo más numerosos posibles, como por ej. la renta de un gran número de familias. Así llamamos, **POBLACIÓN, COLECTIVO O UNIVERSO** el conjunto de individuos o elementos (ya sean personas, animales o cosas) definido sin ambigüedad, sobre los que se va a realizar la investigación estadística.

- **Elementos de una población:** cada una de las unidades o componentes de la población se denomina elemento o unidad estadística de la población. Evidentemente el "elemento" es el objeto de observación estadística, y debe estar definido con absoluta precisión, evitando todo posible tipo de ambigüedad o confusión, tanto para los agentes entrevistadores, como para los usuarios de las estadísticas. Si las unidades estadísticas no están bien definidas carece de sentido la suma de los datos, pues sumaremos unidades heterogéneas, y las medidas cuantitativas pierden todo su valor informativo. Los elementos o unidades estadísticas pueden ser de cualquier naturaleza: una persona, una familia, un animal, una casa. Por otra parte, los elementos de una población pueden ser simples (un individuo) o compuestos (una familia).
- **Tamaño de la población:** el número total de elementos que integran una población recibe el nombre de tamaño de la población, pudiendo ser ésta finita o infinita, según que el número de elementos que la componen sea un número finito o infinito, por ejemplo el conjunto de funcionarios de un Ministerio sería una población finita, en cambio el conjunto de piezas producidas por una máquina en un proceso continuo de producción sería una población infinita.
- **Caracteres de una población:** los elementos de una población poseen una serie de cualidades, propiedades o rasgos comunes que se denomina en Estadística caracteres o características. Por ej. si tenemos una población formada por los funcionarios de un determinado Ministerio, todos los elementos (funcionarios) poseen los siguientes caracteres: nivel de estudios, edad, retribución, antigüedad en la función pública, etc. De ello se deduce que no se pueden observar todos los caracteres de los elementos de una población y es importante el proceso de selección de los mismos que

van a someterse análisis para el éxito de la investigación estadística.

Cada uno de los caracteres de los elementos de una población puede presentar 2 ó más situaciones diferentes posibles, que reciben el nombre de modalidades. Las modalidades deben de ser incompatibles y exhaustivas, de modo que cada elemento de la población presente una, y solamente una, de las modalidades del carácter. Los caracteres más sencillos admiten solamente dos modalidades, es decir, son caracteres dicotómicos, por ej. Sexo (Hombre, Mujer), Fuma (Si, No). El número de modalidades de un carácter variará en función de la información de que se disponga.

1.2.2. Variable. Tipos de variables. Escalas de medida.

Para poder responder correctamente a muchas de nuestras preguntas, lo primero que tenemos que saber es qué tratamiento se le debe dar a los datos. Porque si te fijas, no todos los datos que podemos obtener son del mismo tipo, por ejemplo, pensemos sobre las respuestas a tres de las preguntas de la posible encuesta

1. La respuesta a la pregunta sexo (hombre o mujer)
2. La respuesta a la pregunta número de hermanos
3. La respuesta a la pregunta altura

Lo primero que podemos observar es que la respuesta a la pregunta 1 no es numérica mientras que las de las preguntas dos y tres sí lo son. La característica que corresponde a la respuesta de la pregunta 1 se llama cualitativa mientras que las variables correspondientes a las preguntas 2 y 3 se llaman cuantitativas. Es claro que las variables cuantitativas permiten realizar cálculos que no podemos hacer con las variables cualitativas. A las distintas posibilidades de la característica se les llama modalidades en el caso cualitativo y valores en el caso cuantitativo. Vamos a ver ahora qué diferencias podemos encontrar entre las variables 2 y 3, porque es algo más complicada. La variable número de hermanos toma valores numéricos que podríamos llamar "aislados", 0,1,2,3,..., pero no puede tomar cualquier valor entre ellos, es decir, no puede tomar el valor 3.5 por ejemplo. Sin embargo, con la altura no ocurre esto. En realidad, la altura puede tomar cualquier valor dentro de unos límites, podemos medirla con tanta precisión como queramos. Podríamos decir que la variable altura puede tomar todos los valores posibles dentro de un intervalo. Así, a la variable que resulta en el caso 2 se le llama variable discreta y a la que resulta del caso 3 se le llama variable continua.

En términos estadísticos:

Los caracteres de los elementos de la población pueden ser:

- Caracteres cualitativos o atributos.
- Caracteres cuantitativos o variables (discretas y continuas).

Caracteres cualitativos o atributos, son aquellos que por su propia naturaleza no se pueden cuantificar y se describen mediante palabras, por ej el sexo, nacionalidad, estado civil, nivel de estudios, etc.

En muchos casos, y con el fin de simplificar el tratamiento de estas variables cualitativas o atributos, se suelen codificar, es decir, se le asigna a cada modalidad un número, sin que esto represente ninguna cuantificación ni ordenación posible.

Caracteres cuantitativos o variables, son aquellos que se pueden describir mediante números, es decir, los que son susceptibles de cuantificación o de medida. Por ej, el salario, la edad, peso, etc.

Dicho de otra forma, un carácter es cuantitativo, si sus diversas modalidades son medibles o numerables, es decir, si a cada una de las modalidades se les asigna un número, que recibe el nombre de *valor* de la variable estadística, resultando, pues, que las diferentes modalidades de un carácter cuantitativo son los diferentes valores posibles o diferentes grupos de valores posibles de la variable estadística.

Dentro de los caracteres cuantitativos o variables pueden considerarse dos *tipos* o clases:

- variables discretas.
- Variables continuas.

Una variable estadística es *discreta* si toma un número finito o infinito numerable de valores, esto significa que entre dos valores puede tomar a lo sumo un número finito de valores, por ej. número de hijos, número de obreros en una fábrica, etc.

Una variable estadística es *continua* si toma un número infinito de valores en un intervalo, por ej. peso, estatura, diámetro de una pieza, etc.

La medición u observación de la variable cuantitativa o cualitativa se puede realizar de acuerdo a diferentes reglas, lo que da lugar a distintas escalas de medida. La *escala de medida* es la referencia numérica o no que se utiliza para medir el valor u observar la categoría que toma la variable en cada elemento. Las principales son:

- Escala nominal: una variable se mide en escala nominal cuando clasifica a los elementos del colectivo estudiado en categorías no numéricas mutuamente excluyentes, entre las cuales no se puede establecer ninguna relación de orden, no pudiéndose fijar, por tanto, ningún origen que sirva de referencia. Pertenecen a esta categoría las observaciones correspondientes, por ej. a las diferentes ramas de actividad económica, estado civil, sexo, ideología política, etc.
- Escala ordinal: las medidas en escala ordinal son aquellas que participando de las propiedades de la escala nominal, se diferencian de éstas en que sí se pueden establecer ordenaciones naturales de sus categorías, existiendo pues algún origen de referencia para tal ordenación. Las observaciones que se pueden obtener sobre niveles de estudios (primarios, medios, superiores y otros), estratificaciones de familias por su capacidad de consumo (bajo, medio, alto), etc, pertenecen a este tipo de escala ordinal.
- Escala de intervalos: una variable se mide en escala de intervalo cuando, además de poseer las propiedades de la escala ordinal puede establecerse de antemano algún tipo de unidad de medida, pudiéndose cuantificar numéricamente las distancias o diferencias entre dos observaciones cualesquiera. Ésta es una escala verdaderamente cuantitativa, encontrándose numerosos ejemplos de ella en Economía, tales como datos referidos a salarios, presupuestos, gastos, pasivos financieros, volúmenes de ventas, etc.
- Escala de proporción o razón: una variable se mide en escala de proporción, cuando además de poseer las propiedades de la escala de intervalo puede determinarse la existencia de un origen de referencia natural y no arbitrario, como puede ocurrir con la edad de los individuos.

Ejercicio: Indica si las siguientes variables son cualitativas o cuantitativas, y en caso de ser cuantitativas, si son discretas o continuas:

1. Número de nacidos en un día
2. Grupo sanguíneo de una persona
3. Tiempo que se necesita para resolver un problema
4. Número de preguntas de un examen
5. Temperatura de una persona
6. Partido político votado en las últimas elecciones
7. Número de goles marcados por un jugador en una temporada
8. Clasificar una pieza como aceptable o deteriorada.

9. Número de veces que una persona utiliza el transporte público al día.
10. Edad de una persona.
11. Estatura del encuestado.
12. Número de coches fabricados mensualmente
13. Presión de los neumáticos
14. Cantidades de mercancías vendidas
15. Que periódico lee
16. Número de cigarrillos que fume diariamente

1.2.3. - Tipos de datos

Por datos se entiende el conjunto de observaciones o mediciones de la variable.

Según la relación que estas observaciones o mediciones tengan con el tiempo, los datos se denominan de sección cruzada, serie de tiempo y de panel.

Los *datos de sección cruzada* son los resultantes de medir u observar la variable en los diferentes elementos del colectivo en un mismo instante temporal. (ej. ventas de cada Departamento El Corte Inglés un día concreto).

Serie temporal es el conjunto de observaciones de una variable ordenados según transcurren en el tiempo. (ej. ventas entre 2000-2008)

Un panel de datos es una combinación de serie de tiempo y datos de sección cruzada.

1.3. - Muestra

No siempre es factible estudiar todos y cada uno de los elementos de la población ya sea por razones de coste, de rapidez en la obtención de la información, o porque los elementos se destruyen en el propio proceso de investigación, por lo que es necesario acudir a examinar sólo una parte de esta población, que se denomina *muestra*. Por tanto, una muestra será todo subconjunto representativo de la población.

En el caso de que nuestra investigación se dirija a toda la población se dice que se realiza un *censo*, mientras que la recogida de la información muestral, se llama *encuesta*. A veces se emplea la expresión encuesta exhaustiva como sinónimo de Censo.

1.4. - Estadística Descriptiva e Inferencia Estadística.

La estadística puede dividirse en dos grandes ramas, perfectamente diferenciadas no solamente por los objetivos que se persiguen, sino también por los métodos que se utilizan. Éstas son:

- La Estadística Descriptiva o Deductiva.
- La Inferencia Estadística o Inductiva.

La *Estadística Descriptiva* tiene por objeto recoger, describir y analizar las características de una población o muestra, tratando de poner de manifiesto la estructura y regularidades existentes en los elementos de la población o muestra, no ocupándose de comportamientos individuales. Pero la estadística descriptiva no sólo se limita a recoger, describir y analizar los datos, sino que también se ocupa de la predicción, es decir, determina el valor de una variable en función del valor que toma otra relacionada con la primera. Esta posible relación se determina mediante el conocimiento de la estructura de los caracteres que estamos investigando.

Si la observación de la población no es exhaustiva y se realiza mediante una muestra, tenemos que recurrir a un proceso de inducción para generalizar los resultados obtenidos de la muestra a toda la población, llegando así a la *Inferencia Estadística* o Estadística Inductiva. Definimos, entonces, la Inferencia Estadística como la parte de la Estadística que se ocupa de generalizar los resultados y conclusiones obtenidos a partir de muestras a toda la población: es decir, infiere las características de la población a partir de muestras.

1.5.- Fases de un estudio estadístico.

Toda investigación en la que interviene la estadística exige, para alcanzar los objetivos prefijados, seguir un itinerario, que constará de más o menos pasos según el objetivo que se persigue.

1.- Delimitación del objetivo del estudio: antes de cualquier aplicación de los métodos estadísticos es fundamental definir claramente el objetivo que se persigue. Esta fase condiciona de forma significativa las decisiones posteriores y de ahí su importancia.

2.- Especificación de las características: deben especificarse cuál o cuáles serán las variables de interés que deben medirse u observarse.

3.- Identificación del colectivo: esta fase del proceso consiste en identificar, sin ambigüedad, los elementos del colectivo objeto de la investigación.

4.- Recogida de datos. Métodos de recogida: los datos son el resultado de medir u observar la variable en los diferentes elementos del colectivo. Una vez que tanto el colectivo como la característica o características a estudiar están claramente definidas, el paso siguiente es proceder a la recogida de información.

Básicamente, existen dos métodos de recogida:

- *Método indirecto.* Se caracteriza porque los datos que se necesitan ya han sido recogidos anteriormente en algún estudio previo, aunque con objetivos no necesariamente coincidentes con el de la investigación en curso. Por ejemplo acudir al Censo para datos relativos a la edad. Este método ahorra generalmente costes importantes y cada día es más frecuente su uso.
- *Método directo.* Con este método la información es recogida por el investigador a partir, bien de la observación exhaustiva de todos los elementos de la población en estudio, bien observando un subconjunto representativo de esa población (muestra).

5.- Ordenación de los datos: una vez que los datos han sido recogidos, e independientemente del método seguido para ello, deben ser revisados a fin de detectar posibles errores o incoherencias. Si la información es poco manejable se recurre a tablas o cuadros que permitan ordenar los datos, es lo que se conoce en estadística como tabulación de los datos.

6.- Representación gráfica: los datos numéricos, por muy organizados que estén, pueden no permitir una transmisión rápida y eficaz de las principales características de información contenida en las tablas o cuadros. Por ello como complemento a la tabulación, se utilizan las imágenes o gráficos que hacen más rápida y accesible la información contenida en los datos.

7.- Análisis e interpretación de resultados: ni la presentación ordenada de los datos obtenida a partir de la clasificación en una tabla, ni la transmisión rápida y eficaz que se consigue con la representación gráfica, permiten agotar las conclusiones de cualquier estudio. Hay que analizar e interpretar esa información, y el objeto de esta asignatura es describir herramientas que sirvan para ello.

Dado que el objetivo es diferente en cada caso, las fases o pasos para alcanzarlo serán también diferentes. Dichas fases serán el resultado de aplicar los procedimientos estadísticos adecuados.