

Curso Básico de Estadística

Alfredo Sánchez Alberca (asalber@ceu.es)



CEU

*Universidad
San Pablo*

©Copyleft

Curso básico de estadística

Alfredo Sánchez Alberca (asalber@gmail.com).

Esta obra está bajo una licencia Reconocimiento-No comercial--

Compartir bajo la misma licencia 2.5 España de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/byncsa/2.5/es/> o envíe una carta a Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Con esta licencia eres libre de:

- ▶ Copiar, distribuir y mostrar este trabajo.
- ▶ Realizar modificaciones de este trabajo.

Bajo las siguientes condiciones:



Reconocimiento. Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciador (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).



No comercial. No puede utilizar esta obra para fines comerciales.



Compartir bajo la misma licencia. Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

- ▶ Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.
- ▶ Estas condiciones pueden no aplicarse si se obtiene el permiso del titular de los derechos de autor.
- ▶ Nada en esta licencia menoscaba o restringe los derechos morales del autor.

Contenidos

1. Introducción a la Estadística
2. Fases del análisis estadístico
3. Estadística Descriptiva
4. Regresión y Correlación

Introducción a la Estadística

1. Introducción a la Estadística

1.1 La estadística como herramienta científica

1.2 Población y muestra

1.3 Muestreo

1.4 Variables estadísticas

¿Qué es la estadística?

Definición (Estadística)

La *estadística* es una rama de las matemáticas que se encarga de la recogida, análisis e interpretación de datos.

La estadística es imprescindible en cualquier disciplina científica o técnica donde se manejen datos, especialmente si son grandes volúmenes de datos, como por ejemplo, la física, la química, la medicina y las ciencias biosanitarias, pero también en la economía, la psicología o las ciencias sociales.

Pero,

¿Por qué es necesaria la estadística?

La Estadística está por todas partes en el mundo que nos rodea y la usamos en multitud de ocasiones en nuestra vida cotidiana aunque no seamos conscientes de ello. Cuando miramos una factura de la luz, o el gasto del teléfono móvil, cuando comprobamos la popularidad de nuestros mensajes en las redes sociales, cuando vemos la posición que ocupa nuestro equipo favorito en la clasificación de la liga, o cuando vemos las predicciones meteorológicas, estamos usando la Estadística. Pero formalmente, podríamos definir la Estadística como una rama de las matemáticas que se encarga de la recogida, análisis e interpretación de datos.

Como tiene que ver con el tratamiento de datos, esto la hace imprescindible en cualquier disciplina científica o técnica donde se manejen datos, especialmente si son grandes volúmenes de datos, como por ejemplo, la física, la química, la medicina y las ciencias biosanitarias, pero también en la economía, la psicología o las ciencias sociales.

Pero, en realidad, ¿porqué es necesaria la estadística?

La variabilidad de nuestro mundo

El científico trata de estudiar el mundo que le rodea; un mundo que está lleno de variaciones que dificultan la determinación del comportamiento de las cosas.

¡La variabilidad del mundo real es el origen de la estadística!

La estadística actúa como disciplina puente entre la realidad del mundo y los modelos matemáticos que tratan de explicarla, proporcionando una metodología para evaluar las discrepancias entre la realidad y los modelos teóricos.

Esto la convierte en una herramienta indispensable en las ciencias aplicadas que requieran el análisis de datos y el diseño de experimentos.

Bien, podemos decir que los científicos tratamos de estudiar el mundo que nos rodea, para comprenderlo y controlarlo. Pero vivimos en un mundo con una que está lleno de variaciones que dificultan la determinación del comportamiento de las cosas. No hay dos personas iguales, ni dos animales iguales, ni dos plantas iguales, ni siquiera existen dos gotas de agua iguales. Esta enorme biodiversidad, que es la mayor riqueza de nuestro mundo, es al mismo tiempo el principal obstáculo para comprender los fenómenos naturales.

Así, cuando por ejemplo estamos estudiando una enfermedad y su cura, es posible que un tratamiento funcione con una persona, pero no sirva a otras, o incluso que deje de funcionar con esa misma persona transcurrido un tiempo, porque la personas también cambiamos con el tiempo.

Podemos decir, por tanto, que *¡La variabilidad del mundo real es el origen de la estadística!*

Así, la Estadística juega un papel clave en el Método Científico, ya que hace de puente entre la realidad del mundo y los modelos matemáticos que tratan de explicarla, proporcionando una metodología para evaluar las discrepancias entre la realidad y los modelos teóricos. Y esto la convierte en una herramienta indispensable en las ciencias aplicadas que requieran el análisis de datos y el diseño de experimentos.

Población estadística

Definición (Población)

Una *población* es un conjunto de elementos definido por una o más características que tienen todos los elementos, y sólo ellos. Cada elemento de la población se llama *individuo*.

Definición (Tamaño poblacional)

El número de individuos de una población se conoce como *tamaño poblacional* y se representa como N .

A veces, no todos los elementos de la población están accesibles para su estudio. Entonces se distingue entre:

Población Teórica: Conjunto de elementos a los que se quiere extrapolar los resultados del estudio.

Población Estudiada: Conjunto de elementos realmente accesibles en el estudio.

A la hora de estudiar un fenómeno natural, lo primero es identificar a las unidades experimentales, a las que les afecta este fenómeno, ya sean seres vivos o no. Así, en el estudio de una enfermedad, las unidades experimentales serían las personas, pero en un proceso de fabricación de pastillas, las unidades experimentales serían las pastillas.

En Estadística, se llama población al conjunto de elementos definido por una o más características que tienen todos los elementos, y sólo ellos. Y a cada elemento de la población se llama *individuo*, aunque un individuo no necesariamente tiene que ser una persona, sino que puede ser cualquier objeto.

Por otro lado, el número de individuos de una población se conoce como *tamaño poblacional* y lo representaremos con la letra N mayúscula.

Por ejemplo, en unas elecciones generales a la presidencia del gobierno, la población serían todos los individuos del estado con derecho a voto. En el estudio de una enfermedad, la población sería todas las personas que tienen la enfermedad. Y en un proceso de control de calidad en la fabricación de un fármaco, la población estaría formada por todos los fármacos que se producen en la fábrica.

A veces, no todos los elementos de la población están accesibles para su estudio. Entonces se distingue entre:

Población Teórica: Conjunto de elementos a los que se quiere extrapolar los resultados del estudio.

Población Estudiada: Conjunto de elementos realmente accesibles en el estudio.

Por ejemplo, en el caso del estudio de una enfermedad, la población teórica sería todas las personas que contraigan la enfermedad, incluso si aún no han nacido, mientras que la población estudiada se limitaría al número de personas enfermas que realmente podemos estudiar (obsérvese que incluso quedarían fuera las personas enfermas actualmente que no supieran que lo están o a las que no pudiesemos acceder físicamente para estudiarlas.)

Inconvenientes en el estudio de la población

El científico estudia un determinado fenómeno en una población para comprenderlo, obtener conocimiento sobre el mismo, y así poder controlarlo.

Pero, para tener un conocimiento completo de la población *es necesario estudiar todos los individuos de la misma*.

Sin embargo, esto no siempre es posible por distintos motivos:

- ▶ El tamaño de la población es infinito, o bien es finito pero demasiado grande.
- ▶ Las pruebas a que se someten los individuos son destructivas.
- ▶ El coste, tanto de dinero como de tiempo, que supondría estudiar a todos los individuos es excesivo.

Como hemos dicho, el científico suele estudiar un determinado fenómeno en una población para comprenderlo, obtener conocimiento sobre el mismo, y así poder controlarlo.

Pero, para tener un conocimiento completo de la población *es necesario estudiar todos los individuos de la misma*.

Sin embargo, esto no siempre es posible por distintos motivos:

- En primer lugar, el tamaño de la población suele ser demasiado grande para poder estudiar uno por uno a todos los individuos. A veces incluso podemos tener poblaciones de tamaño infinito, como por ejemplo en el estudio de una enfermedad, ya que también formarían parte de la población las personas que aún no han nacido pero que acaben desarrollándola.
- Otras veces, las pruebas a que se someten los individuos para su estudio son destructivas. Por ejemplo en un proceso de control de calidad en la fabricación de un fármaco, el fármaco suele someterse a pruebas que lo alteran, por lo que ya no se podría comercializar. En este caso, un estudio completo de la población, acabaría con la población y no podríamos vender ni un sólo fármaco.
- Finalmente, a veces el problema es simplemente una cuestión de que el coste, tanto de dinero como de tiempo, que supondría estudiar a todos los individuos es excesivo.

En cualquier caso, es raro que podamos estudiar a todos los individuos de una población para tener un conocimiento completo de la misma.

Muestra estadística

Cuando no es posible o conveniente estudiar todos los individuos de la población, se estudia sólo una parte de la misma.

Definición (Muestra)

Una *muestra* es un subconjunto de la población.

Definición (Tamaño muestral)

Al número de individuos que componen la muestra se le llama *tamaño muestral* y se representa por n .

Habitualmente, el estudio de una población se realiza a partir de muestras extraídas de dicha población.

Generalmente, el estudio de la muestra sólo aporta conocimiento aproximado de la población. Pero en muchos casos es *suficiente*.

Así pues, cuando no es posible o conveniente estudiar todos los individuos de la población, no queda más remedio que estudiar sólo una parte de la misma. Esta pequeña parte de la población que se estudia se conoce como muestra.

En Estadística, se llama muestra a cualquier subconjunto de la población, y al número de individuos que componen la muestra se le llama tamaño muestral y lo representaremos por n minúscula, para distinguirlo del tamaño de la población que habíamos representado con N mayúscula.

Habitualmente, el estudio de una población se realiza a partir de muestras extraídas de dicha población, pero debemos ser conscientes que el estudio de las muestra sólo aporta un conocimiento aproximado de la población. Es decir, nunca vamos a tener un conocimiento completo de la población a partir de las muestras, sin embargo, si la muestra es lo suficientemente grande y representativa de la población, podremos lograr tener un conocimiento de la población suficiente, para comprender el fenómeno e incluso tomar decisiones sin asumir demasiados riesgos.

Determinación del tamaño muestral

Una de las preguntas más interesantes que surge inmediatamente es:

¿cuántos individuos es necesario tomar en la muestra para tener un conocimiento aproximado pero suficiente de la población?

La respuesta depende de varios factores, como la variabilidad de la población o la fiabilidad deseada para las extrapolaciones que se hagan hacia la población. Por desgracia no se podrá responder hasta casi el final del curso.

En general, cuantos más individuos haya en la muestra, más fiables serán las conclusiones sobre la población, pero también será más lento y costoso el estudio.

En el momento de tomar una muestra para estudiar una población, una de las preguntas que surge de manera inmediata es: *¿cuántos individuos es necesario tomar en la muestra para tener un conocimiento aproximado pero suficiente de la población?*

La respuesta no es sencilla, y depende de varios factores, como puede ser la variabilidad de la población o cómo de preciso o aproximado queremos que sea nuestro conocimiento de la población. Parece lógico pensar que en una población con gran variabilidad habrá que estudiar más individuos que en una población con poca variabilidad para tener un conocimiento aproximado de esta. Del mismo modo, si queremos un conocimiento muy aproximado a la realidad, tendremos que estudiar más individuos que si no necesitamos un conocimiento tan preciso.

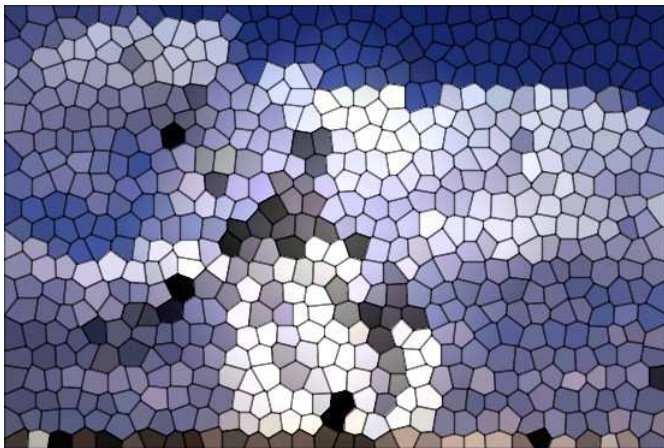
En general, cuantos más individuos haya en la muestra, más preciso será nuestro conocimiento de la población y más fiables serán las conclusiones sobre esta, pero también será más lento y costoso el estudio. Así pues, antes de comenzar cualquier estudio de una población debemos plantearnos siempre estas preguntas y determinar el tamaño muestral necesario, y no más, para tener un conocimiento suficiente de la población.

Por desgracia no podremos responder de manera operativa a esta pregunta este curso, ya que se requieren conocimientos avanzados que se verán en la segunda parte del curso.

Determinación del tamaño muestral

Muestra pequeña de los píxeles de una imagen

¿De qué imagen se trata?



¡Con una muestra pequeña es difícil averiguar el contenido de la imagen!

Para entender a qué nos referimos cuando hablamos de un tamaño muestral suficiente para comprender lo que ocurre en la población, podemos utilizar el siguiente símil en que se trata de comprender el motivo que representa una fotografía.

Como sabéis, una fotografía digital está formada por multitud de pequeños puntitos llamados pixels que se disponen en una enorme tabla de filas y columnas (cuantas más filas y columnas haya se habla de que la foto tiene más resolución). Aquí la población estaría formada por todos y cada uno de los pixel que forman la foto. Por otro lado cada pixel tiene un color y es la variedad de colores a lo largo de los pixels la que permite formar la imagen de la fotografía.

Supongamos que queremos comprender el motivo de la fotografía y para ello tomamos sólo una pequeña muestra de los pixels de ella, como esta que aparece en la pantalla. ¿Serías capaz de averiguar de qué se trata?

Seguramente no has podido averiguar el motivo de la fotografía, porque en este caso el número de pixels que hemos tomado en la muestra es insuficiente para comprender toda la variabilidad de colores que hay en la foto.

Determinación del tamaño muestral

Muestra mayor de los píxeles de una imagen

¿De qué imagen se trata?



¡Con una muestra mayor es más fácil averiguar el contenido de la imagen!

Sin embargo, si tomamos una muestra mayor, como esta otra, seguramente que ya si sabrías decir cuál es la imagen de la foto. ¿Eres capaz?

Determinación del tamaño muestral

Población completa de los píxeles de una imagen

Y aquí está la población completa



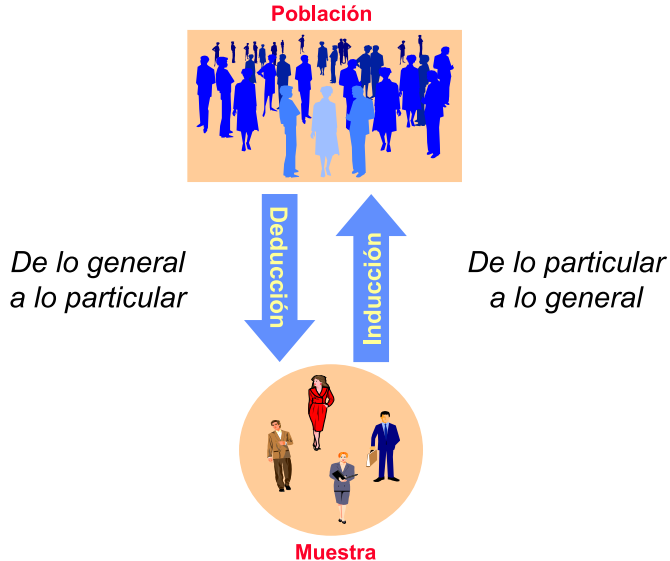
¡No es necesario conocer todos los píxeles para averiguar la imagen!

Efectivamente, se trata de unos molinos de viento, y si has sido capaz de averiguarlo es porque en la segunda muestra el número de pixels tomados en la muestra era suficiente para comprender el motivo de la fotografía.

Evidentemente, cuanto mayor sea la variabilidad de colores de la fotografía, mayor será el tamaño muestral requerido para comprender el motivo de la foto, y cuanto menos variabilidad de colores haya en la foto, menos pixels habrá que tomar, hasta el punto de que en una foto donde no hubiese variabilidad de colores, es decir, donde todos los pixels tuviesen el mismo color, bastaría con tomar un pixel para conocer el motivo de la foto.

Bien, pues esto mismo pasa en las poblaciones que estudiaremos en el ámbito de las Ciencias de la Salud.

Tipos de razonamiento



Como se ha dicho antes, habitualmente realizaremos el estudio de la población a partir de muestras y luego trataremos de extrapolar lo observado en la muestra al resto de la población. A este tipo de razonamiento que saca conclusiones desde la muestra hacia la población se le conoce como razonamiento inductivo.

A diferencia del razonamiento deductivo que va de lo general a lo particular, o en nuestro caso de la población a la muestra, el razonamiento inductivo no garantiza la certeza de las conclusiones, por lo que debemos ser cuidadosos a la hora de generalizar sobre la población lo observado en la muestra, ya que si no se hacen bien las cosas podrían cometerse errores.

Seguramente muchos habréis oído hablar de aquel científico loco que no sabía porqué la gente se emborrachaba y decidió experimentar. Un día probó whisky con leche, y se emborrachó, otro día probó ron con leche y se emborrachó, y otro día probó ginebra con leche y también se emborrachó. ¿Sabéis cuál es la conclusión a la que llegó? Efectivamente, concluyó en su estudio que la causa de la embriaguez era la leche.

En este caso, el científico loco aplicó bien el método inductivo, sin embargo, las conclusiones fueron falsas porque no hizo un buen diseño del experimento y porque la muestra de bebidas tomadas no era representativa. Si hubiese mezclado la leche con agua, habría visto enseguida que la leche no era la causa de su embriaguez.

Esto mismo podría ocurrir en nuestros estudios, si no tenemos cuidado a la hora de diseñar el experimento, y sobre todo, si no tomamos una muestra representativa de la población. Cuando hablamos de que una muestra sea representativa, nos referimos a que refleje las mismas características de la población de la que se ha tomado, pero a pequeña escala. Más adelante veremos cómo debemos tomar la muestra para garantizar su representatividad.

Tipos de razonamiento

Características de la deducción: Si las premisas son ciertas, garantiza la certeza de las conclusiones (es decir, si algo se cumple en la población, también se cumple en la muestra). Sin embargo, *¡no aporta conocimiento nuevo!*

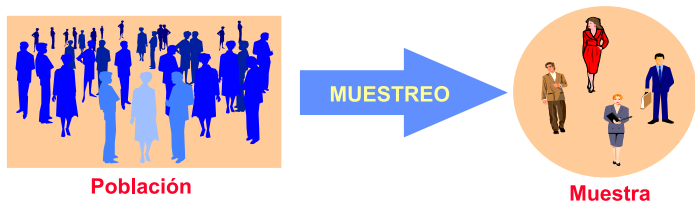
Características de la inducción: No garantiza la certeza de las conclusiones (si algo se cumple en la muestra, puede que no se cumpla en la población, así que ¡cuidado con las extrapolaciones!), pero *¡es la única forma de generar conocimiento nuevo!*

La estadística se apoya fundamentalmente en el razonamiento inductivo ya que utiliza la información obtenida a partir de muestras para sacar conclusiones sobre las poblaciones.

Muestreo

Definición (Muestreo)

El proceso de selección de los elementos que compondrán una muestra se conoce como *muestreo*.



Para que una muestra refleje información fidedigna sobre la población global debe ser representativa de la misma.

El objetivo es obtener una muestra representativa de la población.

En Estadística, al proceso por el cual se seleccionan a los individuos de la muestra se le llama muestreo.

Y como hemos visto antes, es muy importante que la muestra sea representativa de la población para que podamos sacar conclusiones fiables sobre la población.

Modalidades de muestreo

Existen muchas técnicas de muestreo pero se pueden agrupar en dos categorías:

Muestreo Aleatorio Elección aleatoria de los individuos de la muestra. Todos tienen la misma probabilidad de ser elegidos (*equiprobabilidad*).

Muestreo No Aleatorio: Los individuos se eligen de forma no aleatoria.

Sólo las técnicas aleatorias evitan el sesgo de selección, y por tanto, garantizan la representatividad de la muestra extraída, y en consecuencia la validez de la inferencia.

Las técnicas no aleatorias no sirven para hacer generalizaciones, ya que no garantizan la representatividad de la muestra. Sin embargo, son menos costosas y pueden utilizarse en estudios exploratorios.

En general, existen muchos procedimientos para seleccionar a los individuos de la muestra, pero se suelen dividir en dos tipos:

Muestreo Aleatorio Que se caracteriza por la elección aleatoria de los individuos de la muestra. Es decir los individuos se eligen al azar, lo cual quiere decir que todos los individuos de la población tienen la misma probabilidad de ser elegidos en la muestra.

Muestreo No Aleatorio: Que cuando los individuos se eligen de forma no aleatoria.

Bien, debemos tener presente que sólo las técnicas aleatorias evitan el sesgo de selección, y por tanto, garantizan la representatividad de la muestra extraída, y en consecuencia la validez de la inferencia que hagamos sobre la población. Por tanto siempre que queramos sacar conclusiones fiables sobre la población debemos utilizar muestreos aleatorios.

Las técnicas no aleatorias no sirven para hacer generalizaciones, ya que no garantizan la representatividad de la muestra. Sin embargo, son menos costosas y pueden utilizarse en estudios exploratorios. En muchos estudios no tendremos capacidad para hacer una elección aleatoria de los individuos de la muestra y nos tendremos que conformar con una muestra dada. En estos casos no deberíamos hacer generalizaciones sobre la población y deberíamos ser muy cautelosos a la hora de sacar conclusiones sobre esta.

Muestreo aleatorio simple

Dentro de las modalidades de muestreo aleatorio, el tipo más conocido es el *muestreo aleatorio simple*, caracterizado por:

- ▶ Todos los individuos de la población tienen la misma probabilidad de ser elegidos para la muestra.
- ▶ La selección de individuos es con reemplazamiento (y por tanto no se altera la población de partida).
- ▶ Las sucesivas selecciones de un individuo son independientes.

La única forma de realizar un muestreo aleatorio es asignar un número a cada individuo de la población (*censo*) y realizar un sorteo aleatorio.

La técnica de muestreo aleatorio más conocida es el muestreo aleatorio simple, que se caracteriza por tres propiedades:

- Todos los individuos de la población tienen la misma probabilidad de ser elegidos para la muestra. Si no no sería aleatorio.
- La selección de individuos es con reemplazamiento, es decir, tras seleccionar a un individuo se devuelve a la población antes de realizar la siguiente elección. Esto tiene el riesgo de incluir al mismo individuo varias veces en la muestra, aunque en la práctica, con poblaciones grandes es casi imposible, pero a cambio se consigue que no se altere la población de partida.
- Las sucesivas selecciones de un individuo son independientes, y por tanto la elección de un determinado individuo no influye en la elección o no de otros.

En general, existen bastantes dudas en la comunidad científica sobre lo que es un muestreo aleatorio y lo que no lo es. Por ejemplo, mucha gente piensa que hacer una encuesta a pie de calle es un muestreo aleatorio, pero en realidad no lo es porque no todas las personas de la población tienen la misma probabilidad de ser encuestadas (hay gente que ni siquiera sale a la calle). La única manera de garantizar que un muestreo es aleatorio es disponiendo de un censo de la población de manera que todos y cada uno de los individuos tengan un número identificativo único y realizando un sorteo, como el de la lotería, para extraer los identificadores de los individuos que formarán parte de la muestra. Este sorteo se suele hacer habitualmente mediante algún programa informático que genere números de forma aleatoria.

Variables estadísticas y atributos

La característica objeto de estudio puede ser de dos tipos:

Atributos: De carácter cualitativo.

Variables estadísticas: De carácter cuantitativo.

A su vez, los atributos se dividen en:

Nominales: No existe un orden entre las modalidades.

Ejemplo: El color de ojos o de pelo.

Ordinales: Existe un orden entre las modalidades.

Ejemplo: El grado de gravedad de un paciente o la calificación de un curso.

Y las variables estadísticas en:

Discretas: Reciben valores aislados.

Ejemplo: El número de hijos o el número de coches.

Continuas: Pueden recibir cualquier valor de un intervalo.

Ejemplo: El peso o la estatura.

Todo estudio estadístico comienza por la identificación de las características que interesa estudiar en la población y que se medirán en los individuos de la muestra. Estas características pueden ser de dos tipos según sean de naturaleza cuantitativa o cualitativa, es decir, si miden cantidades o cualidades. Las características de naturaleza cualitativa se conocen como atributos o variables cualitativas, mientras que las de naturaleza cuantitativas se conocen como variables estadísticas o cuantitativas.

Los atributos, a su vez, pueden ser de dos tipos, nominales, cuando no existe un orden natural entre las modalidades o valores que puede tomar el atributo, como por ejemplo el color de ojos, que puede ser negro, azul, verde, etc. pero no tiene sentido ordenar los colores; u ordinales, cuando existe un orden natural entre las modalidades, como por ejemplo el grado de gravedad de un paciente que puede ir de leve, moderado, grave o muy grave siguiendo un orden de menor a mayor gravedad.

Por su parte las variables estadísticas también pueden clasificarse como discretas, cuando toman valores aislados que suelen ser números enteros, como por ejemplo el número de hijos de una familia, que puede ser 0, 1, 2, 3, es decir, un número entero positivo pero no puede tomar valores decimales como 1.27; o continuas cuando si pueden tomar cualquier valor dentro de un intervalo real, como por ejemplo la estatura, donde, entre dos estaturas cualesquiera como 1.60 y 1.70 podemos encontrar infinitas estaturas posibles.

Conocer el tipo de las variables u atributos es fundamental pues las técnicas que se utilizarán para describirlas dependen del tipo de característica.

A la hora de seleccionar las variables que se estudiarán conviene tomar todas las variables que puedan tener relación con el fenómeno que se pretende estudiar, y en esto cabe más pecar por exceso que por defecto, ya que si con el trascurso del estudio se observa que una variable no aporta nada, siempre se puede descartar a posteriori, mientras que si se observa que falta una variable importante, en la mayoría de los casos no se podrá volver atrás para medirla.

En el caso de las variables numéricas, es importante también explicitar las unidades en que van a medirse.

La matriz de datos

Las variables o atributos a estudiar se medirán en cada uno de los individuos de la muestra, obteniendo un conjunto de datos que suele organizarse en forma de matriz que se conoce como **matriz de datos**.

En esta matriz cada columna contiene la información de una variable y cada fila la información de un individuo.

Ejemplo

	Edad (años)	Sexo	Peso (Kg)	Altura (cm)
José Luis Martínez	18	H	85	179
Rosa Díaz	32	M	65	173
Javier García	24	H	71	181
Carmen López	35	M	65	170
Marisa López	46	M	51	158
Antonio Ruiz	68	H	66	174

Una vez seleccionadas las variables, estas se medirán en cada uno de los individuos de la muestra con lo que se obtendrá un conjunto de datos que se organiza en forma de matriz conocida como matriz de datos. Es importante organizar la matriz de manera que en cada columna contenga la información de una variable y cada fila la información de un individuo.

En este ejemplo puede observarse la matriz de datos correspondiente a una muestra de cinco individuos en los que se han medido 4 características, tres variables, que son la Edad, el Peso y la Altura, y un atributo que es el Sexo. Como puede apreciarse la primera columna contiene las edades de todos los individuos de la muestra, la segunda el sexo, la tercera los pesos y la última las alturas, mientras la información de cada fila hace referencia al mismo individuo, de manera que la primera fila, por ejemplo, contiene los datos de José Luis Martínez que tiene 18 años, es hombre, pesa 85 Kg y mide 179 cm.

Fases del análisis estadístico

Normalmente un estudio estadístico pasa por las siguientes etapas:

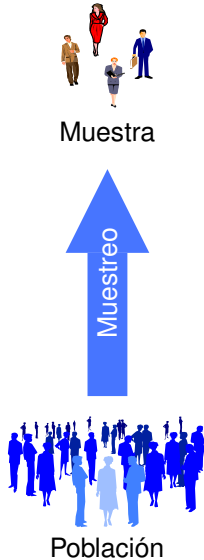
1. El estudio comienza por el diseño previo del mismo en el que se establezcan los objetivos del mismo, se determinen de las variables que se medirán en los individuos de la población y cuál será el tamaño muestral requerido.
2. A continuación se seleccionará una muestra representativa del tamaño establecido. De esto se encarga el **muestreo**.
3. El siguiente paso consiste en estudiar las muestras extraídas y obtener resultados numéricos que resuman la información contenida en las mismas. De esto se encarga la **estadística descriptiva**.
4. La información obtenida es proyectada sobre un modelo matemático que intenta reflejar el comportamiento de la población. Tras construir el modelo, se realiza una crítica del mismo para validarlo. De todo esto se encarga la **inferencia estadística**.
5. Finalmente, el modelo validado nos permite hacer suposiciones y predicciones sobre la población de partida con cierta confianza.

El ciclo estadístico

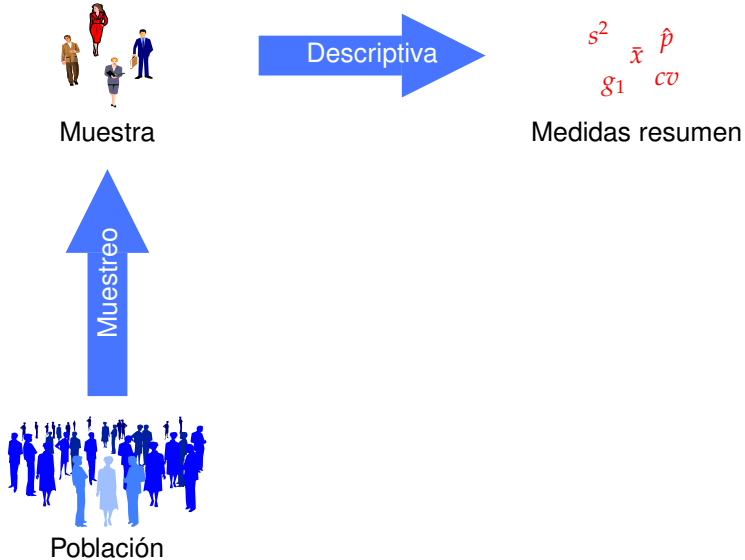


Población

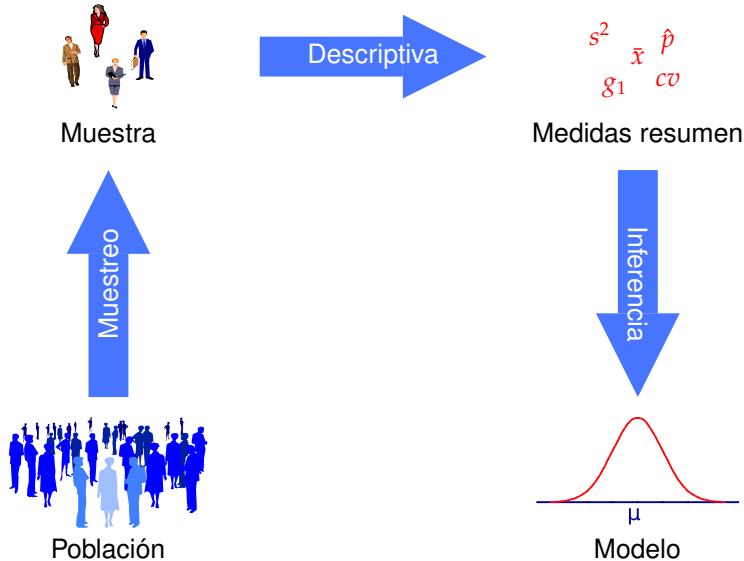
El ciclo estadístico



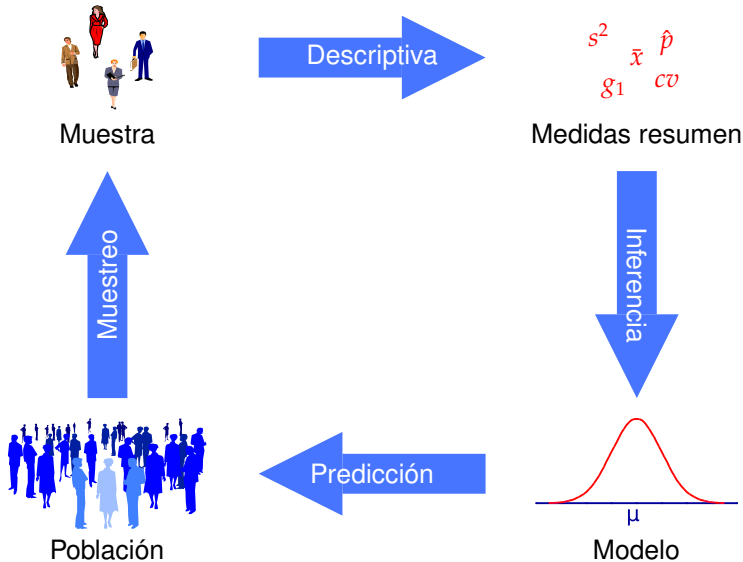
El ciclo estadístico



El ciclo estadístico



El ciclo estadístico



Para acabar con esta introducción a la Estadística, veremos cuáles son las fases por las que suele pasar cualquier estudio estadístico, y las partes de la Estadística que se encargan de ello.

1. En general siempre partiremos de una población en la que queremos estudiar un fenómeno, y se hará un diseño previo del estudio donde se formulen los objetivos del mismo, qué variables se medirán en los individuos de la población, qué precisión queremos para nuestras conclusiones y qué tamaño muestral será necesario para ello.
2. A continuación se debe extraer una muestra representativa de la población y de esto se encarga el **muestreo**.
3. El siguiente paso consiste en estudiar la muestra extraída y presentar de manera ordenada y resumida la información que esta contiene. De esto se encarga la **estadística descriptiva**.
4. A continuación, la información obtenida es proyectada sobre un modelo matemático que intenta reflejar el comportamiento de las variables estudiadas en la población. Tras construir este modelo, se debe realizar un contraste o una crítica del mismo. Si lo que predice el modelo discrepa con lo observado en la muestra, habrá que reformular el modelo o construir otro nuevo, mientras que si encaja en lo observado en la muestra, lo daremos por válido y ello nos permitirá hacer predicciones sobre la población de partida con cierta fiabilidad.

En esta primera parte del curso nos centraremos en la Estadística Descriptiva, es decir, la que se encarga de describir la información contenida en la muestra, pero debemos tener presente que la verdadera potencia de la Estadística está en completar este ciclo y poder llegar a comprender y controlar el fenómeno que afecta a toda la población.