

Introducción a la Estadística

D.S. Gómez-Reverte, M.D. Molina, J. Mulero, M.J. Nueda y A. Pascual

*Depto. de Estadística e Investigación Operativa
Universidad de Alicante*

Gitest

Redest

1

Introducción a la Estadística

1.1 Estadística: definición

El término Estadística tiene dos acepciones fundamentales. Por un lado la Estadística como ciencia o método científico y por otro lado la estadística o estadísticas como conjunto o colecciones de datos. Este segundo concepto es muy usado hoy en día para referirnos a resultados ya elaborados en un estudio en el que se empleó la Estadística como método.

Dado que la Estadística es una disciplina muy amplia, existen diferentes definiciones de la misma según el enfoque en el que se plantee. Para este curso utilizaremos la siguiente definición:

Definición 1.1.

La Estadística es la parte del método científico que mediante el análisis matemático nos permite obtener información sobre la realidad que nos rodea.

La Estadística constituye una poderosa herramienta para generar conocimiento y ha experimentado un vigoroso desarrollo desde sus orígenes hasta nuestros

días.

Actualmente, se aplica en todas las áreas del saber y, de manera determinante, en las Ciencias Sociales. Por ejemplo, en Administración de Empresas se utiliza la estadística para evaluar la aceptación de un producto antes de comercializarlo; en Economía, para medir la evolución de los precios mediante números índice o para estudiar los hábitos de los consumidores; en Sociología para realizar investigación social estudiando los perfiles y dinámica de colectivos sociales; en el ámbito de las Relaciones Laborales elaborando análisis de salarios, paros, accidentes, expedientes de regulación de empleo, etc.; en Criminología para el análisis de los rasgos de la delincuencia así como la prevención del crimen. En general, en las Ciencias Sociales la estadística se utiliza para medir las relaciones entre variables y hacer predicciones sobre ellas.

1.2 Breve introducción histórica

Desde los comienzos de las distintas civilizaciones han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera, huesos, para contar el número de personas, animales o ciertas cosas. Desde que surgen los primeros estados (Babilonios (3000 a.C.), Egipcios (2200 a. C.)), se han recogido datos sobre sus habitantes con el objetivo principal de recaudar impuestos y tributos, y reclutar a jóvenes para el ejército.

Durante los siglos XVII y XVIII los estados europeos comienzan a realizar censos de población y a recopilar de manera sistemática datos demográficos, sociales y económicos. Por tanto, hasta el siglo XIX, la Estadística es una ciencia descriptiva que utiliza medias y gráficos para sintetizar datos sociales y económicos. La necesidad de estimar cantidades desconocidas a partir de muestras va transformando paulatinamente la disciplina en una ciencia normativa para extraer conclusiones de los datos, prever la evolución de las variables y guiar la toma de decisiones en ambiente de incertidumbre; esta transformación es posible por la incorporación del concepto de probabilidad.

En el siglo XIX, Gauss introduce la distribución normal como modelo de los

errores de medida y Quetelet, padre de la sociología cuantitativa, utiliza una distribución para describir y estimar las características sociales medias de los miembros de una comunidad.

A finales de siglo, Francis Galton y Karl Pearson en Inglaterra inventaron métodos para medir relaciones entre variables sociales e introdujeron la idea de regresión y de coeficiente de correlación. En el siglo XX la Estadística se extiende a todos los campos científicos gracias a los importantes avances realizados a principio de siglo en Inglaterra por Ronald A. Fisher, Egon Pearson y Jerzy Neyman. La expansión de sus aplicaciones a todos los campos científicos ha dado lugar a disciplinas específicas como la Econometría, la Biometría o la Psicometría. En la actualidad, la Estadística es probablemente una de las disciplinas científicas más utilizadas y estudiadas en todos los campos del conocimiento humano.

1.3 Etapas de un estudio estadístico

El proceso de investigación supone un conjunto de etapas que podemos representar mediante el siguiente esquema:

1. **Planteamiento del problema:** consiste en definir el objeto de la investigación y precisar el universo o población al que se refiere el estudio.
2. **Planificación del trabajo de campo:** en esta etapa se toman decisiones con respecto a los procedimientos de entrevista, características del muestreo, diseño de herramientas, etc.
3. **Recopilación de información:** incluye la recogida de los datos y también la depuración de la información obtenida, es decir, tratar los problemas de la no-respuesta, los errores de campo, los errores de oficina, los datos desaparecidos y los datos anómalos.
4. **Análisis de los datos:** esta etapa presenta varias fases:
 - a) **Análisis descriptivo:** organización (mediante tablas y gráficos) y resumen (en pocos valores que proporcionen la máxima información po-

sible) los datos disponibles para extraer la información relevante en nuestro estudio.

- b) **Inferencia estadística:** en el caso de trabajar con muestras, la inferencia se encarga de generalizar los resultados de la muestra a la población y obtener conclusiones generales. La inferencia se basa en suponer que la población sigue un modelo o una distribución conocida y los datos que tenemos son realizaciones aleatorias de ese modelo. Para cuantificar la fiabilidad de estos resultados se recurre al **cálculo de probabilidades**.
- c) **Validación del modelo:** finalmente es necesario diagnosticar la validez de los supuestos del modelo que nos han permitido interpretar los datos y llegar a conclusiones sobre la población.

5. **Interpretación:** obtener las conclusiones de los resultados obtenidos.

En el proceso descrito hemos definido las tres ramas fundamentales de la Estadística como ciencia y que son objeto de estudio en las asignaturas habituales de Estadística:

1. Estadística Descriptiva.
2. Cálculo de probabilidades.
3. Inferencia.

En este manual, que está ligado a las asignaturas de Estadística en las Ciencias Sociales, nos vamos a centrar en la Estadística Descriptiva.

1.4 Conceptos básicos

Generalmente el objetivo de un estudio estadístico es conocer información sobre alguna característica de cierto conjunto de elementos. Este conjunto de elementos normalmente tiene un tamaño excesivamente grande para abarcarlo en su totalidad, por lo que nos centraremos en la obtención de información y estudio sobre algunos de ellos, un subconjunto, que podremos hacer extensible al total.

Para matizar este problema, se suelen emplear los siguientes términos:

- **Población:** Conjunto de personas, objetos, ideas o acontecimientos sometido a una observación estadística.
- **Individuo o elemento:** Cada uno de los elementos de la población.
- **Muestra:** Subconjunto de una población.
- **Carácter o variable:** Cada una de las propiedades, rasgos o cualidades que poseen los elementos de una población y que son objeto de estudio. Haremos la siguiente clasificación:
 - **Variable cualitativa o categórica:** Los valores que toman no se pueden cuantificar. Cada uno de estos valores se denomina categoría, clase o modalidad. Pueden ser Ordinales o Nominales dependiendo de si se puede establecer un orden entre las diferentes categorías o no. Ejemplos de este tipo de variables son el carácter rango militar y el sexo, respectivamente.
 - **Variable cuantitativa o medibles:** Los valores que toman se pueden cuantificar o medir. Pueden ser Discretas (los valores que pueden tomar son aislados) o Continuas (pueden tomar cualquier valor de la recta real o de un intervalo). Ejemplos de este tipo de variables son el carácter número de hermanos y el precio de unas acciones respectivamente.
- **Parámetro:** es un valor numérico calculado a partir de una fórmula matemática, obtenida a partir de datos de la población.
- **Estadístico:** es un valor numérico derivada de un conjunto de datos de una muestra, con el objetivo de estimar o inferir características de una población o modelo estadístico. Se trata por tanto de una variable aleatoria que depende de la muestra.

Ejemplo 1.1.

Unos grandes almacenes realiza todos los años una prueba para seleccionar personal. La siguiente tabla recoge la información de 20 candidatos sobre 7 de las variables que aparecen en la ficha de cada uno de ellos.

La variable NOTA es el resultado de la prueba realizada; CALIF es la nota codificada en las categorías suspenso, aprobado, notable o sobresaliente; y CONVOC es el número de veces que ha realizado las pruebas.

Esta nos servirá para analizar los diferentes tipos de variables que aparecen y para aplicar los análisis estadísticos aprendidos en los temas siguientes.

ID.	EDAD	SEXO	NOTA	CALIF.	CONV.	ALT.	PESO
1	18	H	7	NOT	1	1.68	60
2	19	H	3.4	SUS	2	1.80	75
3	17	H	5.3	APR	1	1.71	60
4	19	M	6.1	APR	2	1.56	50
5	22	M	5	APR	3	1.70	57
6	21	H	5	APR	4	1.79	75
7	22	H	9	SOB	1	1.64	58
8	19	M	4.1	SUS	1	1.65	55
9	19	H	3	SUS	4	1.85	80
10	28	H	5	APR	2	1.70	66
11	27	M	5	APR	5	1.75	70
12	22	H	6.3	APR	3	1.65	58
13	23	M	5.3	APR	1	1.80	78
14	18	H	5.5	APR	2	1.62	62
15	18	M	9.5	SOB	1	1.60	64
16	37	M	7.8	NOT	1	1.72	65
17	56	H	8	NOT	1	1.78	80
18	19	H	3.6	SUS	4	1.67	70
19	20	H	4	SUS	1	1.87	90
20	19	H	5	APR	2	1.55	58

Ejemplo 1.1 (continuación).

Veamos como clasificamos cada una de estas variables.

- EDAD, variable cuantitativa discreta.
- SEXO, variable cualitativa nominal.
- NOTA, variable cuantitativa discreta.
- CALIF., variable cualitativa ordinal.
- CONVOC., variable cuantitativa discreta.
- ALTURA, variable cuantitativa continua.
- PESO, variable cuantitativa continua.

Según veremos en el próximo tema algunas variables cuantitativas discretas serán tratadas como continuas, ya que nos será útil trabajar con los posibles resultados agrupados por intervalos.