

Tema 57

Usos de la Estadística: Estadística Descriptiva e Inferencial. Métodos básicos y aplicaciones de cada una de ellas

57.1 Concepto de Estadística. Introducción histórica

Le podemos dar dos significados a la palabra estadística:

- (1) Datos numéricos relativos a un conjunto de elementos o colección de datos numéricos.
- (2) Ciencia que tiene por objeto dar métodos para el tratamiento de las masas de datos de observación y su aplicación.

A las dos acepciones anteriores podemos añadir una tercera debida a Barnett:

“La estadística es la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones que entrañan incertidumbre”.

Etimológicamente el término “estadística” tiene su raíz en la palabra “estadista”, y esta a su vez en el término latín “status”. De aquí nace su primera vocación, la de constituirse como la exteriorización cuantitativa de las cosas del estado.

En este sentido, los antecedentes de la “Estadística” son tan remotos como lo puede ser la historia del hombre. Es fácilmente imaginable que las sociedades humanas más primitivas estuvieran interesadas en enumerar sus característi-

cas más elevantes: familias, hombres aptos para la guerra, utensilios de caza, cabezas de ganado, tec. Las referencias históricas nos proporcionan las primeras evidencias de recuentos, situándolas en el censo del emperador Yao en la China del año 2238 a.C., y en documentos asirios, egipcios y griegos que preceden a los más cercanos del Imperio Romano, en el que la preocupación por la actividad censal de los individuos y bienes del estado tenía una clara finalidad tributaria y militar.

Posteriormente, el avance general del conocimiento cuantitativo de las cosas del estado en sus facetas de recogida de información, descripción y análisis de la misma, adquirió una base más científica a través de las mejoras introducidas por las dos escuelas estadísticas más importantes: la alemana y la inglesa.

Pero en realidad la gran transformación de la Estadística, que la ha convertido en una ciencia susceptible no solamente de discutir la realidad, sino de modelizarla utilizando los métodos del Análisis Matemático, surge precisamente de su vinculación a éste a través del Cálculo de Probabilidades.

El origen del Cálculo de Probabilidades se sitúa en el S. XVII, atribuyéndose a las aportaciones que Pascal realizó sobre algunos problemas clásicos de los juegos de azar. Pero en realidad, ya a partir del S. XV algunos matemáticos notables como Racciolli, Cardano, Kepler y Galileo, habían esbozado unas primeras formalizaciones de algunos esquemas aleatorios.

Esta nueva ciencia fué tomando cuerpo y vinculándose fuertemente a la Teoría de Funciones a lo largo de los siglos XVIII y XIX, y comienzos del XX merced a los logros de figuras tan notables como Bernouilli, Leibniz, Bayes, Laplace, Tchebicheff, Kolmogorof, Markov,... El resultado de todo ello ha sido la construcción de un modelo de comportamiento de los llamados fenómenos estocásticos en el que puede encuadrarse toda experiencia o evidencia empírica que revista carácter de aleatoriedad.

La fusión de estas dos vertientes de mejora del conocimiento: la Estadística como recogida, descripción y análisis de la información, y el Cálculo de Probabilidades, se ha plasmado en una nueva rama floreciente de esta disciplina: la Estadística Matemática, surgida en las primeras décadas de este siglo y cuyo fruto, producto de aportaciones de hombres como Yule, Fisher, Neyman, Pearson,..., ha sido la disponibilidad de eficaces instrumentos que permitan poner en relación los datos recogidos con algún modelo ideal de probabilidad, y ayudan a descubrir en la evidencia empírica algún tipo de regularidad estocástica.

Resumiendo, históricamente, la Estadística ha comenzado por ser descriptiva. Ha sido necesario ante todo acumular información, criticarla, ponerla en condiciones, analizarla y sintetizarla. Posteriormente, después de haberse comprobado analogías, descubierto permanencias estadísticas, reconocido un cierto número de distribuciones tipo, observado algunas formas de dependencia estructurales bastante grandes, la Estadística llegó a ser explicativa, gracias en particular al Cálculo de Probabilidades.

La Estadística, por tanto, 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 ésta incertidumbre pueda ser medida en términos de probabilidad. Por ello la Estadística se preocupa

por los métodos de recogida y descripción de datos, así como de generar técnicas para el análisis de esta información.

A raíz de todo lo expuesto, podemos dividir el estudio de esta disciplina en:

- (1) Estadística Descriptiva
- (2) Cálculo de Probabilidades
- (3) Estadística Teórica o Inferencia Estadística

57.2 Investigaciones estadísticas

Una investigación estadística consta usualmente de cuatro partes: diseño o planificación, recogida de datos, análisis y presentación.

Planificación:

El que la planificación de una investigación es extremadamente importante no necesita ninguna justificación. Una de las experiencias más "duras" del estadístico profesional, al recibir datos de un cliente es el verse forzado a decir que, debido a una mala planificación, es imposible obtener conclusiones a partir de los datos y que, por tanto, es necesario empezar de nuevo. La planificación es muy importante a causa de un hecho a menudo descuidado: cuando planificamos la investigación, el investigador puede afectar a la elección del modelo teórico para los datos obtenidos. Si la planificación es buena, es posible usar un modelo que conduzca a conclusiones claras, mientras que si la planificación es mala, no podrá construirse ningún modelo.

El tamaño de la muestra no determina si una investigación es buena o mala. Investigaciones bastantes amplias pueden ser inútiles.

Una parte importante de la planificación es formular un plan que sea el mejor posible entre el rango de medios disponibles.

No existe un método de planificación general para el inmenso número de investigaciones que se pueden realizar en la práctica. Sin embargo se aplican ciertas características comunes.

A continuación damos algunos pasos en la planificación que son relevantes en una investigación.

Paso 1: preliminares

- Escribir la información disponible para la investigación.
- Distribución de recursos (tiempo, personal, dinero, etc.)
- Formulación del problema práctico.
- Propiedades deseables de la solución (población estudiada, precisión deseada)
- Clase de investigación.

Paso 2: resultados teóricos

- Construcción del modelo.
- Formulación del problema teórico.
- Propiedades deseables de la solución.

- Construcción del plan para la investigación.
- Elección del método estadístico.

Paso 3: pasos prácticos finales

- Plan detallado para recogida de datos.
- Plan para proceso de datos.
- Plan para análisis estadísticos.
- Plan para la presentación de resultados.

Los pasos 1 y 2 se discuten a menudo al mismo tiempo, pero el 3 debe discutirse después de los anteriores.

Recogida de datos:

Es un término general que puede significar, por ejemplo, la recogida de datos de un experimentador midiendo con un instrumento o de un entrevistador preguntando a la gente sobre sus opiniones.

Análisis:

Puede ser de diversas formas: en casos una condensación o resumen de los datos en una tabla o diagrama; para ello usamos la Estadística Descriptiva. Para un análisis más detallado de investigaciones muestrales se usa alguna clase de análisis estadístico, apoyado en la Estadística Matemática.

Presentación:

Puede consistir en ilustraciones gráficas, un resumen de resultados y conclusiones y recomendaciones prácticas.

57.2.1 Investigación completa e investigación muestral

En una investigación completa se estudia la población entera. Tales investigaciones son importantes en muchas áreas, por ejemplo en estadísticas oficiales de nacimientos y muertes, etc.. Son igualmente importantes en Industria: determinación de la producción total, inspección de cada unidad cuando se transportan productos costosos, etc.

A menudo ocurre que una investigación completa es demasiado costosa o consume mucho tiempo para realizarla, o es imposible (por ejemplo cuando un producto se destruye en su muestreo). Por último, las investigaciones completas no requieren una teoría estadística.

En una investigación muestral es examinada sólo parte de la población. Tales investigaciones tienen amplias aplicaciones, por ejemplo, en sondeos de opinión, experimentos y control estadístico de calidad. Se puede decir que toda actividad experimental consiste en investigaciones muestrales. En efecto, el investigador puede realizar sólo una secuencia limitada de medidas, que se pueden considerar como muestras de una secuencia conceptualmente infinita de observaciones. La

idea de que deben tenerse conjuntos de datos muy grandes para obtener conclusiones dignas de confianza es totalmente incorrecta; a veces, una muestra muy pequeña puede producir información suficiente.

Una investigación muestral es generalmente más imprecisa que una investigación completa. Sin embargo, esto no es siempre así. Cuando los items son examinados por humanos, períodos de fatiga pueden dar lugar a malos resultados; una inspección muestral bien planeada da lugar a un mejor resultado. Además el término investigación muestral dado incluye el experimento más general.

57.2.2 Investigaciones comparativas y no comparativas

El significado de estos términos es obvio. Una Investigación comparativa consiste en una comparación de dos o más poblaciones. Ejemplos: ¿los coches de la marca A consumen más gasolina que los de la marca B?, ¿el medicamento A es mejor que el B?. En una investigación no comparativa sólo se estudia una población. Ejemplos: ¿cuánta gasolina consumen los coches de la marca A?, ¿cuál es el efecto del medicamento A?. La planificación de los dos tipos de investigación es bastante diferente.

57.3 Estadística Descriptiva

La Estadística Descriptiva no es una teoría, sino un modelo. Su fin no es explicar, sino al contrario, describir con los medios apropiados, separar lo esencial, resumirlo y medirlo.

Podemos definir la Estadística Descriptiva de muchas formas, pero consideraremos la definición siguiente:

“La Estadística Descriptiva es un método de descripción numérica de conjuntos numéricos”.

Las unidades sobre las que se va a hacer el estudio deben pertenecer a un grupo homogéneo y bien delimitado y se necesita saber que características del grupo interesan.

La Estadística Descriptiva se aplica a todos los dominios de investigación cuantitativa: investigación demográfica, económica, agronómica, biológica, industrial,... No obstante, los métodos de recogida de información, la crítica de los datos recopilados, y más aún la interpretación de los resultados obtenidos, no son propiamente hablando, de la competencia de la Estadística Descriptiva.

Contnidos del estudio de la Estadística Descriptiva son: medidas de posición, de dispersión, las rectas de regresión, el cálculo de números índices, el estudio de series cronológicas,...

57.4 Inferencia Estadística

Esta parte de la Estadística tiene como objetivo obtener indirectamente información respecto de una población mediante la información directa que nos da una muestra (conjuntos de unidades de la población de las cuales se obtiene la información).

La Inferencia, particularmente la Teoría de Decisiones, ha jugado y juega un papel muy importante en la vida del hombre desde la antigüedad, sobre todo en el ámbito de las Ciencias Experimentales y en económico-empresarial: preocupación de los economistas por futuras exportaciones de ciertos productos, predicción del comportamiento de la bolsa de valores, interés por diseños y resultados de experimentos biológicos, químicos, físicos, etc., son problemas que requieren de la Inferencia Estadística.

La construcción de los modelos probabilísticos presentada en cualquier texto de probabilidad, y base esencial para el estudio de la Inferencia Estadística, sigue el esquema del razonamiento deductivo: se establecen hipótesis respecto al mecanismo generador de los datos y con ellos se deducen las probabilidades de los valores posibles. La Inferencia Estadística sigue el esquema inverso (de razonamiento inductivo): dadas frecuencias observadas de una variable, se infiere el modelo probabilístico que han generado los datos.

Los procedimientos de Inferencia Estadística pueden clasificarse en función del objetivo de la inferencia (métodos paramétricos frente a no paramétricos) y según el tipo de información que utilizan (métodos clásicos frente a bayesianos).

Los métodos paramétricos suponen que los datos provienen de una distribución conocida cuyos parámetros se desea estimar.

Los métodos no paramétricos no suponen conocida la distribución y sólo introducen hipótesis muy generales respecto a ésta, utilizándose especialmente para juzgar que las hipótesis en el enfoque paramétrico no son contradictorias con la muestra.

Los métodos clásicos suponen que los parámetros son cantidades fijas desconocidas y que la única información existente respecto a ellos es la contenida en la muestra.

Los métodos bayesianos consideran los parámetros como variables aleatorias y permiten introducir información a priori sobre ellos.

57.5 Métodos básicos y aplicaciones de la Estadística Descriptiva e inferencial. El aparato matemático de la Estadística Inferencial

La Teoría de la Probabilidad es el aparato matemático usado en la Estadística Inferencial.

Supongamos un fenómeno aleatorio del que efectuamos un cierto número de observaciones que recopilamos en una tabla de modo ordenado.

Queremos saber si el experimento en cuestión se ajusta a algún modelo probabilístico teórico. Cabe pues preguntarse: ¿cómo hay que decidir que el modelo teórico se ajusta a las observaciones que indican nuestros datos?.

Si A es un suceso que puede presentarse o no al realizar el experimento aleatorio, conoceremos (generalmente) su probabilidad según el modelo teórico trazado. Para que el modelo teórico represente un buen ajuste, ha de suceder que en una repetición (en número de veces suficientemente grande) la frecuencia relativa del suceso A coincida aproximadamente con la probabilidad calculada de modo teórico.

Además, si la probabilidad del suceso A es pequeña y la frecuencia relativa $\frac{n}{N}$, para N suficientemente grande, del suceso A no es pequeña, el modelo teórico no define un buen ajuste.

Pero si la probabilidad de A es pequeña y al realizar el experimento una vez se presenta A , se puede “suponer” que el modelo teórico no es el adecuado.

Una vez hayamos decidido el modelo teórico que se ajusta a nuestros datos y obtenidas las características descriptivas (media, varianza, etc.) de los datos recogidos en la tabla, se consideran los datos como valores observados de ciertas variables aleatorias cuya distribución de probabilidad conocemos en todo o en parte. Las características descriptivas de la muestra sirven para hacer estimaciones de las características desconocidas de la población.

Todo esto es lo que constituye el aspecto descriptivo de las aplicaciones del método estadístico.

La Teoría de la Probabilidad suministra medios para un análisis científico de los datos estadísticos.

Otro aspecto concreto es el de la estimación de parámetros. La Teoría de Estimación fué establecida por Fisher y tiene como fin principal, determinar ciertos parámetros de la población mediante valores obtenidos a partir de las muestras. Algunos métodos de estimación son:

(1) El método de la máxima verosimilitud, que es posiblemente el más importante. La idea de este método es la de maximizar cierta función, llamada función de verosimilitud.

(2) Método de los momentos.

(3) Método de los mínimos-cuadrados.

(4) Método de la mínima χ^2 .

...

A veces deseamos utilizar nuestros datos para aceptar o rechazar una hipótesis, relativa a la distribución de probabilidad considerada. Este es otro tipo de problema de decisión.

Un último aspecto de las aplicaciones de la Teoría de la Probabilidad es el de la predicción Si de los datos estadísticos recopilados previamente poseemos información sobre las distribuciones de probabilidad de las variables aleatorias

que intervienen, la Teoría de la Probabilidad suministra soluciones al problema de predicción planteado.

57.6 Aplicaciones

Estadística Descriptiva

Su principal aplicación es “describir” los datos obtenidos en tablas denominadas distribuciones de frecuencias. Una posterior aplicación consiste en la obtención de determinadas características de la población como pueden ser las medidas de centralización, dispersión, simetría o asimetría, aplastamiento, coeficientes de correlación, rectas de regresión, números índices, series temporales, etc.

Al analizar las estaturas de los mozos de reemplazo de un determinado año, para la reducción de todos esos datos y obtención de unos números representativos del conjunto, no necesitamos, en principio, el empleo de la Inferencia Estadística. La Estadística Descriptiva se aplicará para dar respuesta (en este ejemplo) a: la altura media de los mozos, concentración de altura alrededor de la media, ¿hay más mozos en el reemplazo con altura mayor que la media?,...

La media, la desviación típica, y las medidas de simetría nos permiten responder a esa preguntas.

Inferencia Estadística

La aplicación básica es tomar muestras y obtener, a través de éstas, conclusiones de la población de la que se han extraído las muestras.

Otras aplicaciones están basadas en las estimaciones puntual y por intervalos, contraste de hipótesis, etc.