

APROXIMACIÓN A LA ESTADÍSTICA DESDE LAS CIENCIAS SOCIALES.

Carrasco Arroyo, Salvador
Dpto. Economía Aplicada
Universidad de Valencia

1. APROXIMACIÓN A LAS CIENCIAS SOCIALES

El desarrollo de la Ciencia moderna ha supuesto una permanente carrera por fijar unas fronteras claras, explícitas y detectables entre el conocimiento científico y el que no lo es. El conocimiento científico es, en nuestra opinión, una parte de la producción simbólica de una sociedad. Las Ciencias Sociales, históricamente han tenido que “sobredemostrar” con frecuencia su “cientificidad” y es a partir de ese proceso cuando surgen las cuestiones sobre cuáles son las finalidades de las ciencias en general, y de las ciencias sociales en particular, así como si existe un único método científico o si por el contrario éste es particular de cada campo de conocimiento.

Podemos establecer dos grupos de ciencias atendiendo a si su objeto de conocimiento guarda o no alguna relación con la realidad (física o social). De ello surge la dicotomía entre Ciencias Factuales o Empíricas, cuyo objeto de conocimiento es algún segmento de la realidad, y las Ciencias Formales, cuyo campo de actividad son entes formales, ideas, pero en ningún caso cosas o procesos reales.

Puede decirse, que las ciencias fácticas o empíricas están formadas por conjuntos de ideas organizadas sistemáticamente, conjuntos que deben adaptarse lo máximo posible a la realidad que sirve de referencia. En este sentido las ciencias sociales son un subgrupo de ciencias fácticas, aquellas que se ocupan de las parcelas de la realidad relacionadas con el hombre como ser social. Por tanto, su ámbito de referencia es la realidad social en sus distintas manifestaciones, los fenómenos sociales. Así pues, el ámbito de conocimiento de las ciencias sociales es el plano humano-social de la realidad que está formado por personas físicas o jurídicas que interactúan y realizan actividades inmersas en un escenario.

La investigación es una actividad humana orientada a la descripción, comprensión, explicación y transformación de la realidad social a través de un plan de indagación sistemática. La investigación social suele iniciarse con la identificación de un problema y la delimitación de unos objetivos que reflejan las intenciones del investigador.

El problema a investigar puede surgir como fruto del interés o curiosidad del investigador por un fenómeno social observado.

Los problemas de investigación no suelen encontrarse aislados y definidos en la realidad social, por lo que el investigador tiene que a partir de la observación de la realidad identificar los problemas, reformularlos hasta que puedan ser investigados. El investigador selecciona tras consulta del marco teórico existente la metodología más adecuada y lleva a cabo la aplicación del protocolo de investigación a través de las técnicas de obtención de información y recoge los datos necesarios para analizarlos con el fin de: explorar, comprender, interpretar o explicar el fenómeno objeto de estudio.

Los científicos sociales han visto el mundo social desde diferentes perspectivas han utilizado diversos modelos de investigación y han recogido la información a través de múltiples técnicas. Esta diversidad de técnicas obedece a cuestiones planteadas desde la filosofía de la ciencia. Durante el último cuarto de siglo las discusiones metodológicas se han centrado en la crítica de la metodología científica (empírico-analítica) utilizada en las Ciencias Sociales y en las nuevas vías metodológicas que han ido emergiendo. Los cambios y avances que continuamente tienen lugar en las ciencias sociales, obligan a una constante actualización y revisión terminológica, así adoptamos los términos de *empírico-analítica*, *constructivista* y *sociocrítica*. De entre ellas, la metodología *empírico-analítica* suele orientarse al contraste de hipótesis, siendo su centro de interés la construcción de conocimiento teórico para describir, explicar, predecir y en lo posible, controlar los fenómenos sociales. En general, su ámbito de aplicación queda reducido a fenómenos observables susceptibles de medición, control experimental y análisis estadístico.

La observación sistemática, la medida, el experimento y la construcción de teorías son aspectos clave de esta metodología, junto con el papel que tiene la estadística en la contrastación de hipótesis. En esta metodología la cuantificación es una consecuencia directa de la medida, que por definición genera datos cuantitativos y exige un análisis de los mismos utilizando modelos estadísticos, en su pretensión de llegar a generalizaciones de carácter probabilístico.

2. ECONOMÍA Y ESTADÍSTICA.

La Economía es probablemente una de las más relevantes dentro de las ciencias sociales, esta ciencia ha alcanzado mediante el recurso al lenguaje formal y al aparato matemático el nivel de abstracción suficiente para: formular hipótesis concretas y deducir a partir de ellas proposiciones que se consideran contrastables. Ahora bien, esta visión absolutamente complaciente del estado o cualidad científica de la Economía no está, sin embargo, libre de crítica.

Son muchas las definiciones de esta disciplina, pero desde nuestro punto de vista se pueden detectar dos perspectivas con implicaciones distintas: la primera perspectiva tiene que ver con la definición de L.Robbins (Robbins, L. 1935), cuando expresaba que "*en la economía pura examinamos la implicación de la existencia de medios escasos con*

usos alternativos". Una segunda aproximación podría referirse a todas aquellas actividades humanas y relaciones sociales encaminadas a la producción y distribución de bienes y servicios con el fin de alcanzar un cierto nivel de bienestar de los miembros de la sociedad.

La segunda definición de la economía tiene una visión más instrumental ya que. *"El problema económico es sencillamente el proceso de procurar el bienestar material de la sociedad. En términos más sencillos, la economía es el estudio de cómo el hombre se gana su sustento diario"* (Heilbroner, 1988). A su vez, dicho problema se concreta en, sobrevivir y progresar de manera que se mejoren esos niveles de bienestar.

La adopción de esa concepción instrumentalista de la economía va acompañada del supuesto de que lo económico es mensurable y que esta mensurabilidad (cardinal y ordinal) lo es en términos monetarios y a precios de mercado, y por tanto, excluidos de la realidad (y también, lógicamente, del análisis económico) otros modos de evaluar los objetos y fenómenos económicos que no se expresen en dinero y a través de los mecanismos de una economía de mercado. La predicción por lo cuantitativo conduce incluso a establecer la barrera entre lo que es científico y lo que no lo es¹, es decir, entre lo que forma parte de la Economía y lo que queda excluido. Para apreciarlo cabe discernir entre tres usos posibles de que pueden ser objeto las mencionadas técnicas econométricas. A saber: la contrastación empírica de hipótesis y teorías, la correcta cuantificación de las relaciones económicas y, finalmente la predicción cuantitativa, el pronóstico del futuro (Thurow, 1988).

Por otro lado, si disponer de una definición de cualquier disciplina, que resulte a la vez suficiente, concisa y operativa, no es tarea fácil, en el caso de la Estadística la cuestión se complica aún más. Es evidente la importancia que tiene y ha tenido la Estadística en el desarrollo de muchas áreas del saber científico. Esta circunstancia supone una seria dificultad a la hora de acotar las características de la estadística como ciencia independiente. Podemos encontrar aplicaciones de esta disciplina en los más variados campos de la actividad humana (Economía, Medicina, Ingeniería, Física, Química, Psicología, Literatura, Agricultura, etc.). La Estadística ha llegado a ser lo que es gracias al uso que de ella han ido haciendo las diferentes ramas del conocimiento y del saber, convirtiéndose en un instrumento metodológico de las demás ciencias.

Si contemplamos la Estadística desde el punto de vista sustancial vemos que los adjetivos de formal y factual podrían serle aplicados, al ocuparse tanto del mundo de las ideas como de los hechos. De igual forma, desde el punto de vista del método nos enfrentamos con una disyuntiva similar, dado que la disciplina podría ser catalogada como teórica (al emplear algoritmos deductivos) y como experimental (al servirse también de procedimientos inductivos).

¹ Llegando incluso a identificar los métodos cuantitativos con los enunciados de hecho y los cualitativos con los enunciados de valor (Katouzian, 1982).

La Estadística es para Newbold (1998) “...lo que hacen los estadísticos”, no sin apostillar: “...Lo que llama la atención es la diversidad de problemas en los que se ven envueltos los estadísticos..”, señalando algunos de los problemas genéricos en los que se ve envuelto el estadístico: sacar conclusiones de la información numérica, manejar la incertidumbre, muestrear la población, analizar relaciones, predecir, tomar decisiones en entornos de incertidumbre, etc.

Por otro lado, es fácil encontrar definiciones no tan extremas como las de Newbold, definiciones que no están libres de las inclinaciones que los distintos autores sienten hacia las concepciones filosóficas que los inspiran, con lo que las definiciones disponibles están orientadas hacia una u otra de las posibles categorizaciones de la ciencia. Así, hemos optado por ofrecer y comentar sólo algunas, de entre las muchísimas existentes, que, bajo nuestro punto de vista, aclaran algún aspecto destacable de la disciplina.

Como punto de arranque partiremos la definición de Kendall y Stuart (1977), quienes conciben a la Estadística como una rama del método científico al definirla con las siguientes palabras: “La Estadística es la rama del método científico que se ocupa de los datos obtenidos contando o midiendo las propiedades de las poblaciones de los fenómenos naturales”, definición que matiza la realizada por Kendall (1968) en un trabajo anterior, cuando afirmó: “la estadística es la matriz de toda ciencia experimental y, por consiguiente, una rama del método científico, si no el Método Científico por excelencia”.

Ambas definiciones, enfatizan el carácter instrumental de la Estadística en la investigación científica, al menos en lo referente a las ramas experimentales, y ponen de manifiesto el carácter “metodológico” de la Estadística, al ser el instrumento que permite examinar las proposiciones teóricas a la luz de los hechos del mundo observable. En tal sentido debería entenderse la frase de Mood y Gaybill (1978) cuando afirman que la Estadística es “la tecnología del método científico”.

La concepción popular de la Estadística como columnas de datos o gráficos recoge, sin embargo, solamente un aspecto del trabajo estadístico, quizás su aspecto menos formal, consistente en compilar, organizar, presentar y analizar información, generalmente de índole cuantitativo. Esta faceta, primigenia en el desarrollo histórico de la Estadística, suele denominarse Estadística Descriptiva. Ahora bien, los desarrollos más importantes de la Estadística se fundamentan en las deducciones derivadas de la Teoría Matemática de la Probabilidad, la cual parte de los axiomas de Kolmogorov y utiliza resultados y desarrollos de otras ramas matemáticas. Esta parte, denominada Estadística Matemática, comprendería la parte más formal de la Estadística que se fundiría con la Estadística Descriptiva en la denominada Estadística Inferencial, constituida por un cuerpo de técnicas diseñadas para resolver problemas de colectivos a partir del estudio de una fracción o muestra, convenientemente elegida, de los mismos. En este sentido, Fernández de Trocóniz (1982) afirma: “El estudio de las propiedades generales de las probabilidades y sus distribuciones es objeto de la Teoría de las Probabilidades. La Estadística de la Inferencia o Teoría de las Muestras utiliza los medios que la Teoría de las Probabilidades pone a su alcance, y establece las teorías de estimación y contrastes que permiten hacer

inferencias sobre los colectivos a partir de la información contenida en las muestras...". De modo que, a pesar de que el carácter instrumental de la Estadística es muy importante, no debe olvidarse a la Estadística como ciencia formal, dado que es este aspecto el que la dota de sólidos cimientos y posibilita su utilización por otras disciplinas de forma rigurosa.

Por tanto, se puede concluir que el método básico de la Estadística es el propio de las ciencias formales, es decir, el lógico-deductivo; aunque, debido al estrecho vínculo que mantiene con el mundo empírico, la heurística de la disciplina viene determinada en gran medida por la Estadística Aplicada. Es precisamente del contacto con otras ciencias, en su calidad de instrumento de las mismas, de donde surgen con frecuencia innovaciones estadísticas, que una vez desarrolladas, validadas e incorporadas a la Teoría, pueden ser aplicadas en otras disciplinas. En este sentido Box (1990) señala que *"los mayores avances en la Ciencia, y en la Ciencia Estadística en particular, generalmente ocurren como resultado de la interacción entre teoría y práctica"*.

Continuando con la revisión de las definiciones más significativas de la Estadística que se encuentran en la literatura, debemos referirnos ahora a aquéllas que ponen el énfasis en el objeto -más que en método- de la disciplina. El objeto de la Estadística puede ser definido como la elaboración de principios y métodos que contribuyan a la toma de decisiones en ambientes no especificados, o incluso, generalizando algo más, como el estudio de la propia toma de decisiones en situaciones inciertas.

La incertidumbre es consecuencia directa de la información incompleta que caracteriza las situaciones del mundo real, información incompleta que puede provenir de errores de medición, de inaccesibilidad a la información completa o de la práctica imposibilidad de determinar cuales son los factores relevantes que intervienen en cada problema. Por tanto, la obtención de información se convierte en otro de los aspectos destacables de la investigación estadística. En este sentido, destaca la interpretación de DeGroot (1970), cuando afirma que: *"La ciencia estadística se ocupa del desarrollo de teorías y técnicas apropiadas para realizar inferencias bajo condiciones de incertidumbre e ignorancia parcial que necesariamente existen en una amplia gama de actividades. La práctica estadística corriente depende fundamentalmente de la formulación de modelos probabilísticos de diversos sistemas físicos, de métodos de recogida y análisis de datos numéricos y del diseño de experimentos eficientes e informativos"*.

El elemento que posibilita la medición de la incertidumbre es la probabilidad. Este concepto aparece para proporcionar una medida de la incertidumbre que comporta un fenómeno o experimento aleatorio. La probabilidad permite pasar de la descripción a la inferencia al dar una medida de la incertidumbre, pudiendo así fundamentar las decisiones que se toman.

En síntesis, podemos señalar que la Estadística es una ciencia, fundamentada en la Teoría Matemática de la Probabilidad, que se ocupa de elaborar teorías y técnicas destinadas a la recogida, transformación y clasificación de información, y a la toma de decisiones en situaciones de desconocimiento, incertidumbre o ignorancia.

Sin embargo, como decíamos, la Estadística es una materia de suma importancia para el investigador al margen del cual sea su especialidad. Esta relación de vínculo tiene una doble dirección: por una lado, la Estadística es un instrumento básico para la investigación económica y social aplicada, y por otro, las necesidades de la ciencia económica y su teorización han impulsado el desarrollo de numerosas técnicas, modelos y conceptos estadísticos que, posteriormente, se han aplicado en otros campos.

Desde la óptica del economista, la Estadística puede ser utilizada de dos modos, según el tipo de actividad a que se dedique: por un lado, la aplicación estadística es un proceso de investigación científica económica (bien sea ésta teórica o aplicada), y por otra parte, su aplicación en la práctica económica, esto es, prioritariamente en la gestión.

La elaboración de una ciencia económica sería, capacitada para ofrecer explicaciones que contribuyan a mejorar el conocimiento de los fenómenos económicos reales presupone, en primer lugar, contar con un conocimiento suficientemente profundo de esta realidad. Con preferencia, la manifestación de los fenómenos se determina mediante su cuantificación. Es en el primer estadio de la investigación, en la fijación del objeto, cuando la Estadística entra a formar parte del instrumental requerido por el economista investigador. En una fase posterior, una vez existe una teoría formulada, capaz de explicar los hechos, la Estadística proporciona las técnicas que permiten contrastar empíricamente las hipótesis y predicciones que la teoría presupone mediante la Inferencia Estadística.

Esta dependencia del método estadístico por parte de la Economía para establecer sus teorías y validarlas es general de las Ciencias Sociales, ya que estudiando el mundo social o económico, en el que resulta imposible la realización de experimentos cruciales para la consolidación empírica de las teorías, para su falsación o para su establecimiento, el científico social o el economista debe recurrir a la observación más que a la prueba; a la recolección de datos y a su análisis más que a la experimentación controlada; y, en este contexto, necesita de la Estadística para modelizar la realidad que observa y poder comprenderla utilizando métodos estadísticos. Desde esta perspectiva la Estadística cumple dos cometidos importantes: por un lado, suministrar informaciones acerca del mundo real que sean operativas tanto para facilitar la elaboración teórica como para la descripción e interpretación de la realidad en el marco de una teoría; y, por otro, validar las teorías mediante contrastación de las hipótesis.

Los economistas no sólo emplean la Estadística como un instrumento que le sirve para analizar problemas, formarse ideas de la realidad y tomar decisiones. Este empleo cada vez superior de estadísticas por los agentes económicos es lógico, dada la tendencia creciente a basar las decisiones empresariales en fundamentos lo más científicos y objetivos posibles. La Estadística aporta, por un lado, los instrumentos e informaciones adecuados para el conocimiento de la realidad en que operan las empresas y, por otro, dota a sus gestores de instrumentos para la toma de decisiones óptimas en ambientes de incertidumbre.

3. UNA HISTORIA DE LA ESTADÍSTICA.

Para conocer una disciplina es imprescindible indagar sobre su origen y evolución, así se expresa Daniel Peña (1992) cuando dice: *“el conocimiento de la historia de una disciplina es importante, al menos en tres aspectos: el primero, para entender su estado actual de desarrollo y la relación entre sus partes; el segundo para comprender su terminología, ya que el nombre de una técnica o de un método está siempre asociado a sus orígenes históricos; y el tercero, para prever su desarrollo futuro”*.

La Ciencia Estadística nace de la confluencia de dos corrientes conceptuales: La Estadística, como actividad consistente en reunir un conjunto de datos concernientes al conocimiento de los Estados o sociedades humanas, y la Teoría de la Probabilidad. Su consolidación en la disciplina que hoy manejamos esta estructurada en base a tres etapas: en una primera, aparece la "Estadística" que cuantifica los hechos sociales, políticos o económicos; en una segunda, donde se estudia el origen del Cálculo o Teoría de la probabilidad y, finalmente una tercera, que integra ambas disciplinas apareciendo lo que denominaremos más tarde como estadística inferencial. Su evolución transcurre a través de dos periodos cuyo punto de inflexión lo marca el siglo XX. El primero de ellos, abarca desde los orígenes más remotos hasta principios del siglo XX y está caracterizado por las dos primeras etapas, Estadística como Ciencia del Estado y el cálculo de Probabilidades. El segundo, se desarrolla a lo largo del siglo XX consolidando la estadística matemática o teoría de la estadística como la estadística moderna.

3.1. Estadística como Ciencia del Estado

La Estadística entendida como una actividad de recuento o recolección de datos para proporcionar información a los Estados o sociedades humanas, tiene su origen cuando estos se organizaron. Podríamos afirmar que la estadística surgió en épocas muy remotas. Su historia empieza posiblemente en la isla de Cerdeña, donde existen monumentos prehistóricos en cuyas piedras se encontraban muescas que previsiblemente permitían contar el ganado cazado.

Su existencia se debía a las necesidades de los Gobiernos o sociedades por conocer hechos políticos, sociales o económicos: Sus dominios, riqueza, sus producciones, las personas que habitaban en los mismos y en principio con fines económicos, demográficos y fiscales. Aquellas incipientes técnicas se conocen como estadística descriptiva, caracterizada por sus facetas de recogida de información, recuento, descripción y análisis de dicha información.

Las referencias a estas actividades de recogida de datos o de recuento son numerosas y entre ellas se pueden citar algunos ejemplos² tales como:

² Ver Escuder (1987), Gutiérrez (1982), Fernández-Abascal et al. (1994), Peña (1991) entre otros.

- En el mundo antiguo, los gobiernos por motivos militares, fiscales y administrativos ordenan el empadronamiento de su población o censos de producción y de personas.
- En China, Confucio cita la confección de un censo para las producciones agrícola, industrial y comercial, en el año 2238 antes de la Era Cristiana.
- En Egipto, los censos y catastros llegaron a poseer carácter institucional y las estadísticas tuvieron una relevancia social de gran magnitud.
- En Israel, en el libro Números, se habla de un censo de personas realizado por Moisés.
- En Roma, estas recolecciones fueron tan destacables que hasta se implantaron censores. Se pueden citar a Servio Tulio, quien en el año 335 a. de C. instituyó la obligación de acudir a las fiestas de aldeas o ciudades y depositar cada ciudadano una moneda, distinta para hombres, mujeres o niños. El recuento de monedas proporcionaba el número de habitantes de cada clase. Se hicieron también varios censos con fines militares, fiscales y de avituallamiento e incluso, según Tácito, el emperador Augusto realizó un balance de las riquezas del imperio. Este emperador, en el año 577 ordenó un censo de la población por su conexión con la cristiandad.

Por su parte, en la Edad Media merece destacada atención la aparición de lo que posteriormente serán las estadísticas industriales. Los Estados recogieron información sobre la población y riqueza que existían en sus dominios. Estos censos (censos romanos, los inventarios de Carlomagno sobre sus posesiones, etc.) pueden considerarse precedentes de la institucionalización de la recogida de datos demográficos y económicos de los Estados modernos, principalmente por razones fiscales.

Con el nacimiento del Estado Moderno en el renacimiento resurge la elaboración de censos y recolección de datos económicos-sociales con el desarrollo de la Contabilidad, las Matemáticas, las operaciones comerciales y. Así, en el siglo XIII y en la República de Venecia, se realizaron las denominadas *Relaciones*, un conjunto de datos que puede compararse a los actuales Anuarios de Comercio Exterior. También en Italia, Sansovino, Botero y Arity se consideran exponentes de lo que posteriormente será la denominada Aritmética Política. Esta Aritmética Política, que es una concepción de Estadística de carácter eminentemente descriptivo, evolucionó durante los siglos XVII Y XVIII tomando un carácter más cuantitativo.

A principios del siglo XVII, la familia Elzevir, en los Países Bajos, publica, en 60 volúmenes, una especie de enciclopedia de los estados. Dicha "enciclopedia" incluía información sobre la Economía y el Comercio de los Estados y se considera como el primer precedente del *Statesman's Yearbook*. Se da así el primer paso en el cambio de aplicación de la Estadística hacia aspectos más amplios de la sociedad, aquellos relacionados a las actividades económicas y comerciales sin fines tributarios.

En el siglo XVII, con el cambio de mentalidad, abandonando la teología para la explicación de muchos tipos de fenómenos, nacen algunos incipientes conceptos estadísticos básicos basados en tres corrientes. Por un lado, la escuela estadística-descriptiva y la escuela aritmética-política y por otro, el Cálculo de Probabilidad, constituyendo así los antecedentes de los métodos estadísticos modernos. Tales corrientes son la:

1. Escuela estadística-descriptiva o descriptiva-histórica.

En Alemania, el tratado de Westfalia (1648) impone la necesidad de una disciplina práctica y jurídica. Su objetivo era la descripción de los Estados o "Staadenskunde", de ahí el nombre de esta escuela como descriptiva. Algunos autores opinan que fue fundada por Herman Conring (1606-1681); otros que conjuntamente por éste y por Vito de Seckendorff. El tratado de Seckendorff de 1660 es un ejemplo inspirado por esta escuela. Dado que la escuela alemana fue su máximo representante, también se conoce a esta corriente conceptual por este nombre. La labor de Conring fue continuada, en el XVIII, por su discípulo Godofredo Achenwall (1719-1772), quien realizó una descripción más sistemática de las tablas numéricas. Éste asignó el término Estadística³ (1748) a una nueva disciplina contemplada como una ciencia independiente al servicio del Estado y que, más tarde, se consolidó como disciplina universitaria. Por ello, se le atribuye como creador de la denominación "Estadística" y también se conoce a esta escuela como "Escuela Universitaria".

2. Escuela inglesa o aritmética-política.

En Inglaterra, la escuela aritmética-política, además de los objetivos de la escuela alemana se planteó por primera vez la búsqueda de las leyes cuantitativas que rigen los fenómenos objeto de estudio. John Graunt, un exitoso comerciante y William Petty, médico, ambos ingleses, pueden considerarse como fundadores de esta escuela. La procedencia de sus creadores da lugar a una de sus denominaciones.

A principios del siglo XVII, Inglaterra fue asolada por una de las más graves pestes (1603) dando lugar a recuentos semanales de bautismos y entierros, los cuales fueron más regulares que los previos. Así, una recopilación de datos sistematizados estaba disponible desde comienzos de dicho siglo para ser analizados. John Graunt (1620-1674) fue la primera persona que realizó inferencias a partir de datos demográficos, en concreto, llevó a cabo estimaciones de la población de Londres y otros análisis demográficos relevantes que se encuentran recopilados en su obra *Natural and Political Observations mentioned in a following Index, and made upon the Bills of Mortality*, publicada en 1662. Sus análisis están desvinculados del concepto de probabilidad.

Por su parte, William Petty (1623-1687) realizó una revisión de la obra de Graunt que publicó en un artículo "*Review of Graunt (1662)*" en 1666. Siguió desarrollando sus investigaciones en la misma línea que Graunt, escribiendo de manera amplia - principalmente ensayos- sobre lo que denominó "Política Aritmética", de ahí que se le atribuya como fundador de esta escuela. Este autor sugirió a la Nación la creación de una oficina central de estadística que recogiese datos sobre la misma.

Debido a los trabajos de Graunt y Petty, se elaboran los primeros censos de población oficiales, tanto a finales del siglo XVII como durante el siglo XVIII. En Europa, el

³ No obstante, dicho término había sido, previamente utilizado por el historiador G. Ghiline en 1589, según Gutiérrez (1982).

primer censo se realizó en Irlanda en 1703 y, en España, las primeras manifestaciones, propiamente dichas, suelen situarse a mediados del siglo XVIII con el Catastro realizado por el Marqués de la Ensenada de 1748 ó con el Censo de 1787 impulsado por el Conde de Florindablanca. A comienzos del siglo XIX, la mayoría de los países europeos recogen información oficial mediante censos de datos económicos, demográficos, climáticos, etc. Paralelamente, surgen las Agencias Oficiales de Estadística y en 1834 se creó en Londres la Royal Society, seguida en 1839 de la American Statistical Association.

3.2. Cálculo de probabilidades.

Esta corriente se desarrolló, en principio, de manera independiente a las dos vertientes conceptuales anteriormente citadas -la escuela descriptiva y la aritmético política-. El antecedente más lejano en el tiempo se sitúa en los juegos de azar. Se sabe que desde tiempos muy antiguos, el hombre ha jugado con instrumentos primitivos -tales como tabas, astrágalos y dados-, utilizando también mecanismos de azar con fines adivinatorios Sin embargo, sorprendentemente, no parece que los griegos, los romanos o los pueblos de la Europa medieval llegaran a tener un claro concepto de las leyes del azar. Adicionalmente, como viene a decir Hacking (1995), se advierten dos aspectos distintos en las anticipaciones de la teoría de las probabilidades: Los problemas combinatorios y los problemas sobre juegos repetidos.

Parece ser que la combinatoria elemental era conocida por los árabes y por los matemáticos del Renacimiento, pero como rama del álgebra más que en su contexto probabilístico. Sin embargo, el concepto de azar era bastante familiar, especialmente en el juego, que se extendió considerablemente, a pesar de los constantes obstáculos puestos por la Iglesia y el Estado. Era lógico que surgieran algunas ideas primitivas acerca de la frecuencia relativa de un acontecimiento, pero la teoría del azar tardó extraordinariamente en aparecer. Los primeros pasos en esa dirección se dieron en los siglos XV y XVI. Italianos como Cardano, Galilei, Paccioli o Tartaglia comienzan a desarrollar conceptos tales como el experimento aleatorio, espacio muestral, regularidad estadística, etc.

El famoso matemático, médico, físico, astrólogo y, además, jugador, Gerolamo Cardano (1501-1576), de origen italiano, fue el primero en dejar un manuscrito en el que se formula explícitamente el concepto de leyes de azar. En su obra *Liber de Ludo* (que no fue impresa hasta 1663) abordó diversos problemas combinatorios, destacando en ella un primer acercamiento a la ley de los grandes números. Utilizó también, en algún problema, la regla de la multiplicación para probabilidades, aunque sólo para el caso de sucesos independientes y la regla de la adición, para sucesos disjuntos. El problema de los puntos fue tratado, junto a Cardano, por otros autores como Galilei, Tartaglia y Fray Luca Paccioli.

Por su parte, el matemático, físico y astrónomo, Galileo Galilei (1564-1642), mediante la observación y el estudio de la naturaleza, hace progresos en la Teoría de la Probabilidad y dejó un fragmento que demuestra que comprendió claramente, el método de calcular cómo se podría tener éxito en el lanzamiento de unos dados.

La opinión generalizada propone que el germen de una primera teoría matemática de las probabilidades habría que buscarlas en la sociedad francesa de mediados del siglo XVII, en la correspondencia entre los matemáticos Pascal (1623-1662) y Pierre de Fermat (1601-1665). El origen de estas investigaciones se encuentra en un problema de juego planteado por un jugador francés, el caballero de Meré, a Pascal en 1654. Se trata de repartir el dinero apostado en un juego de azar interrumpido antes de la victoria de uno de los jugadores, es decir, se trata de medir lo que cada jugador puede esperar del azar en las partidas futuras. La solución de Pascal, y sobre todo la de Fermat, se relaciona con el análisis combinatorio, cuyas primeras herramientas estaban en vía de elaboración. En la correspondencia que sostuvieron, aparece el nombre de geometría del azar para este tipo de problemas, así como el concepto de esperanza matemática.

Simultáneamente, un nuevo escenario: Holanda y un nuevo contribuidor: Huygens. El matemático holandés Christian Huygens (1620-1699) escribió un libro sobre juegos de azar. Esta obra -publicada en latín *Ratiociniis in a Jeeae Judo, en Exercitiorum Mathematicorum* y que data de 1657 -se considera el primer libro impreso sobre probabilidad (Seneta. 1983). Según Hacking (1995), este autor, en una visita a París, escuchó las soluciones planteadas por Pascal y Fermat que eran públicamente difundidas y conocidas en el entorno de la época; otros autores dicen que se basó en la difusión de su correspondencia. En cualquier caso, en los intercambios de ideas entre ambos personajes se basó para realizar su obra, en la que incluyó aportaciones propias de incipientes conceptos tales como esperanza matemática, vida probable, vida media y muestreo con o sin reemplazamiento. Aplicó la Teoría de Probabilidad a la demografía, elaborando una curva de mortalidad bajo la suposición de una tasa de mortalidad uniforme, a partir de las cifras de Graunt, para aplicarla al cálculo de anualidades.

A mediados del siglo XVII surge la figura del matemático suizo Jakob Bernoulli (1654-1705). Éste, gracias a su formación filosófica junto a su gran intuición matemática, pudo enfrentarse a los problemas conceptuales que aparecen en las aplicaciones del Cálculo de Probabilidades. Su obra clave, el gran tratado sobre Probabilidad, *Arts Conjectand⁴* consta de cuatro partes. Las tres primeras están en la línea de sus antecesores, principalmente en la de Huygens, aunque también desarrolló nuevos métodos combinatorios y aplicó los mismos a nuevos problemas. Pero es en la parte cuarta donde expuso una teoría que revoluciona el Cálculo de Probabilidades en dos sentidos: Aportando, por un lado, la concepción subjetivista de la probabilidad -como el grado de creencia que se tiene de que ocurra un suceso- y, por otro lado, la primera demostración rigurosa de la ley débil de los grandes números, que denominó Teorema Áureo.

Paralelamente a estos trabajos, el francés emigrado a Londres, Abraham de Moivre (1667-1754) publica en Gran Bretaña varias ediciones de su obra *Doctrine of Chances* en 1756. En sus tratados, este matemático-físico, definió conceptos tales como

⁴ El Arte de la Conjetura obra póstuma publicada por su sobrino Nicolás Bemouilli (1687-1759).

la probabilidad, la esperanza matemática, la independencia de sucesos, la probabilidad condicionada y se establece la regla de adición.

Durante el siglo XVIII el cálculo de probabilidades se extiende a problemas físicos y actuariales (seguros marítimos). El factor principal, impulsor de su desarrollo durante este período, fue el conjunto de problemas de astronomía y física, que surgieron ligados a la contrastación empírica de la teoría del matemático, filósofo y astrónomo inglés Sir Isaac Newton (1642-1727). A lo largo del siglo XVIII y parte del siglo XIX, la investigación en física y astronomía estaba dirigida por el paradigma de Newton. Estas investigaciones iban a ser de importancia fundamental en el desarrollo de la estadística como bien señala Peña (1983) "... Bernoulli, Simpson, Lagrange, Laplace y Gauss realizan sus aportaciones probando y contrastando el paradigma de Newton, principalmente en Astronomía".

A principios del XVIII en el Reino Unido, la figura más controvertida de este período: Thomas Bayes (1702-1761) con su famoso artículo *An Essay towards Solving a Problem in the Doctrine of Chances* -publicado por su amigo Richard Price en 1763 tras su muerte- ha sido objeto de debates sobre los fundamentos de la Inferencia Estadística. En este artículo se aporta un enfoque probabilístico del razonamiento inductivo, usando las probabilidades observadas de acontecimientos para comparar cómo de admisibles son las hipótesis que podrían explicarlos. En el teorema que lleva su nombre se basará la futura corriente bayesiana de inferencia estadística.

Pero, las contribuciones señaladas hasta ahora sólo suponen el inicio del Cálculo de Probabilidades, pues es a partir del siglo XIX cuando comienza una nueva etapa que servirá para consolidar los conceptos y llegar al estado actual de esta materia. Durante el siglo XIX aparecen dos corrientes importantes constituidas por los matemáticos franceses y la escuela rusa:

Corriente francesa

Se abre con una obra capital, considerada la mayor contribución de esta época, *Theorie Analytique des Probabilités* del matemático Pierre Simon de Laplace (1749-1827), publicada en 1812. En ella se encuentra la llamada definición "clásica" de probabilidad - como "el cociente entre el número de casos favorables al suceso y el total de casos posibles, supuestos igualmente verosímiles"; concepto que permaneció hasta principios de este siglo. Además, analizó el poder matemático de transformaciones como la función generatriz de momentos, propuso la función característica y aportó una primera versión del teorema central del límite para una suma de variables aleatorias independientes (que demostrará formalmente Liapunov en 1901). También, desarrolló la ley normal para describir la variabilidad de los errores de medida, planteó el problema de predecir los valores de una variable a partir de los de otras relacionadas con ella y formuló y estimó el primer modelo estadístico explicativo.

Pertenecientes a este mismo período y miembros de la misma escuela francesa son Simeón D. Poisson, A. Cauchy, J. Bienaymé, A. Cournot y J. Bertrand. Los trabajos del matemático Simeón Denis Poisson (1781-1840) son, en gran parte, continuación de los de

J. Bernouilli, De Moivre y Laplace. Aportó el modelo de distribución de probabilidad que lleva su nombre -Distribución de Poisson- y su relación con el modelo binomial. Además, aportó una versión del teorema central del límite para una suma de variables independientes no idénticamente distribuidas. Destaca su obra *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités* (1837) en la que aborda problemas de contexto social.

Por su parte, el matemático francés y barón Agustin-Louis Cauchy (1789-1857) demostró el teorema central, de forma rigurosa, en el contexto de los mínimos cuadrados y haciendo uso de la función característica. Otras contribuciones son las de Jules de Bienaymé (1796-1878), estadístico y matemático francés, que aportó la desigualdad de Bienaymé-Chebyshev; Antoine Cournot (1801-1877), matemático, economista y filósofo francés, ofreció una visión frecuencia lista de la probabilidad y Joseph Bertrand (1822-1900), matemático francés, que escribió un manual de texto tal y como actualmente se concibe.

Corriente rusa

Durante el siglo XIX la otra corriente fue representada por la escuela rusa, integrada por un conjunto de matemáticos que dieron un fuerte impulso al Cálculo de Probabilidades. Buniakovsky (1846) y Ostrogradsky -formados en París y en contacto con Laplace, Cauchy y Poisson- realizaron las aportaciones iniciales. Motivado por ellos, Chebyshev funda la Escuela de San Petersburgo. Éste último, en su obra *Des Valeurs Moyennes* (1867) presentó la desigualdad que lleva su nombre (Chebyshev-Bienaymé) y demostró formalmente la ley de los grandes números de Poisson. Son discípulos suyos Markov, que contribuyó decisivamente a la teoría de los procesos estocásticos, y Liapunov, que logró demostrar de forma rigurosa el teorema central del límite para la suma de variables aleatorias independientes no idénticamente distribuidas, en condiciones generales.

Con el siglo XX llega la axiomatización del concepto de probabilidad, basada en el espacio probabilístico, a partir de lo cual se definen las variables aleatorias como funciones medibles sobre el espacio de sucesos. Sergei Bernstein, Richard Von Mises y Emil Borel destacan en la primera etapa. Sin embargo, será la axiomatización realizada por A. N. Kolmogorov (1903-1987), en su obra *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933), la actualmente aceptada. Este enfoque se basa en la Teoría de la Medida, desarrollada previamente por E. Borel (1871-1956) y H. Lebesgue (1875-1941).

3.3. La Estadística Matemática

La confluencia entre la Estadística, en su sentido tradicional, y la Teoría de la Probabilidad surgió la Estadística Moderna o Estadística Matemática. A este respecto, se considera que el primer acercamiento entre sus creadores fue realizado por el

matemático belga Adolphe Quetelet⁵(1796-1874), a quien Schumpeter (1971) considera como el fundador de la Estadística Moderna. Sus trabajos, *La Methode Statistique* (1848) y *Sur l'Homme et le développement de ses facultés* (1835) representan los primeros pasos hacia la aplicación de la probabilidad para la medición de la incertidumbre en las Ciencias Sociales. Utilizó la distribución normal con datos de tipo social y biológico, aportando el concepto de "hombre medio" y de "ajuste de distribuciones". A pesar de tal contribución, la fusión entre ambas es aún muy precaria y serán las aportaciones del mundo anglosajón las que sientan las bases de la Estadística Moderna. Al igual que en el desarrollo y consolidación de la Teoría de la Probabilidad fue decisiva la influencia de Newton, en los comienzos de la Estadística Matemática son las teorías de C. R. Darwin (1809-1882) las que jugaron un papel decisivo. Así se aprecia como, en las últimas décadas del siglo XIX, aparece un conjunto de trabajos que se encuadran dentro de esta corriente investigadora tales como los estudios de Galton, Edgeworth y Pearson. Autores como Stigler (1986) opinan que la labor conjunta de estas figuras, que destacan -por separado- en otros campos del conocimiento: antropología, economía y filosofía de la ciencia, sienta las bases de la Estadística actual.

Con Francis Galton (1822-1911) se inicia, de manera definitiva, la utilización de la Estadística en las ciencias experimentales. Fue el primero en resaltar la necesidad de acudir a métodos estadísticos para contrastar la teoría de Darwin. En su trabajo principal *Natural Inheritance* -publicado en 1889- estudia exhaustivamente la distribución normal y ofrece dos aportaciones fundamentales: la regresión y la teoría de la correlación. Fue quien introdujo el concepto de línea de regresión, la técnica que utilizó para realizar la comparación de las estaturas de padres e hijos. También ocasionó la entrada del uso de los percentiles en la Estadística. Galton fue un hombre de profunda curiosidad intelectual que contribuyó notablemente en los estudios de herencia, de la psicología, de la antropología, de la meteorología y de la estadística. La importancia de Galton radica no sólo en su nuevo enfoque del problema de la dependencia estadística, sino también en su influencia directa sobre Edgeworth, Weldon y Pearson, entre otros. El primer departamento de Estadística, en el sentido actual de la palabra, fue patrocinado por Galton y llevó su nombre. Asimismo, la fundación de la revista *Biométrica* fue posible gracias a su generoso apoyo económico.

En esta fase de apogeo de la Estadística Matemática, no se puede olvidar la figura de Francis Ysidro Edgeworth (1845-1926). Este economista irlandés, a raíz de su amistad con Galton, intuyó el potencial de la obra de éste último y apreció la utilidad de la estadística en la investigación social y económica. En *Correlated Averages* (1892) aplica y generaliza los conceptos de correlación y regresión. Sus trabajos en múltiples campos (curvas de frecuencia, análisis de la varianza, números índices, series temporales, etc.) le han valido una merecida posición en la historia de esta ciencia.

⁵ Aunque existieron precedentes de esta fusión en los trabajos de Cotes (1722), Mayer (1750), Boscov (1755) en Astronomía, de Simpson (1756) en la distribución de los errores, o los de Gauss (1755) y Legendre (1805) en relación al método de los mínimos cuadrados en problemas de relaciones lineales.

Aunque los trabajos de Galton tuvieron influencia sobre Edgeworth, fue el zoólogo Walter Weldon (1860-1906) quien los aceptó plenamente y comenzó estudiar la aplicación de los métodos estadísticos a la biología animal. Para ello, buscó la colaboración del matemático y filósofo Karl Pearson, quien realizó una labor esencial para la difusión de los nuevos métodos. Sin duda, la gran figura de estos tiempos es Karl Pearson (1857-1936) cuyas principales contribuciones a la Estadística de la época fueron el estudio de las distribuciones bidimensionales, los trabajos dedicados al análisis de la regresión y correlación y el uso de la distribución ji-cuadrado para pruebas de bondad de ajuste. Además, Pearson dedicó una gran parte de sus esfuerzos a la publicación de las tablas estadísticas que permitieron la utilización práctica de los nuevos métodos, con lo que contribuyó decisivamente a su rápida difusión.

Discípulos de Pearson fueron Yule y Gosset: A George Udni Yule (1871-1951) se debe el inicio del análisis multivariante, la distribución binomial negativa y sus estudios sobre series temporales y procesos estocásticos. William Sealy Gosset (1876-1937) -más conocido con el seudónimo de "Student"- se interesó por la distribución exacta de los estadísticos en el muestreo para el caso de muestras pequeñas gracias a los obstáculos de tipo económico que suponían la aplicación de los métodos desarrollados por Pearson para el caso de muestras grandes. Para abordar este problema, aplicó, por primera vez, el Método de Montecarlo simulando el proceso de obtener muestras de una distribución normal, obteniendo la distribución "t" para el estudio de muestras pequeñas. Dicha distribución se publicó con el nombre de "Distribución t de Student". Fuera de Inglaterra también se producen otros avances: En Francia por parte de Borel, Fréchet, Poincaré; en Italia por Gini y en Rusia por Chuprov y Romanovsky.

En las primeras décadas del siglo actual se estaba creando una disciplina. Como indica Stigler (1986), hasta 1900 muchas mentes trabajaban por separado en el campo de la estadística, mientras que a partir, aproximadamente, de esa fecha se puede identificar a un nutrido grupo de estadísticos trabajando sobre la materia estadística. Puede decirse que en los primeros años de este siglo, la Estadística, en su concepción moderna, había comenzado su andadura. En la década de los años veinte se consolida lo que hoy se conoce como Inferencia Estadística Clásica. Los trabajos de Fisher, Neyman y Pearson configuran la teoría de estimación por intervalos y de verificación de hipótesis.

A partir de 1912 y durante aproximadamente una década, los avances de la Estadística fueron, en gran medida, monopolizados por Fisher, el discípulo de Pearson de más renombre. Sus contribuciones han sido numerosas y de gran importancia. La época de la Teoría de las Distribuciones Muestrales comienza con Sir Ronald Aylemer Fisher (1890-1962), que logró sistematizar las ideas sobre la teoría de la estimación, introduciendo conceptos tales como estimador, eficiencia en la estimación y puso de manifiesto la aplicabilidad del método de máxima verosimilitud en la construcción de estimadores adecuados. Sus aportaciones a la teoría de la estimación y verificación, al diseño de experimentos, a la distribución del coeficiente de correlación, los coeficientes de regresión, de correlación múltiple y del estadístico F para muestras procedentes de poblaciones normales, al análisis de la varianza y a los métodos multivariantes, lo convierten en una figura clave. Peña (1983) considera que en su obra fundamental,

Statistical Methods for Research Workers (1925), aparece el cuerpo metodológico básico de lo que constituye la Estadística actual. Esto es, el problema de elegir un modelo a partir de datos empíricos, la deducción matemática de sus propiedades, la estimación de los parámetros y la validez final del modelo mediante un contraste de hipótesis (Peña, 1983). Otras figuras relevantes durante esta época son los discípulos y colegas de Fisher, Bartlett y Cochran. También destacan Finney, Irwin y Yates.

Los otros pilares básicos de la Estadística Teórica actual fueron Jerzey Neyman (1894-1982) y Egon Sharpe Pearson (1895-1980), quienes, en un conjunto de artículos publicados en torno a los años cuarenta, clarifican la teoría de los contrastes estadísticos - introducen la idea de hipótesis alternativa, marcando las diferencias entre la estimación de parámetros y el contraste de hipótesis-, separan los tipos de error, definen la potencia de un test y proponen el test de la razón de verosimilitud.

Todos estos avances producidos en los años 30, 40 Y 50 establecen el núcleo de la hoy llamada Inferencia Estadística Clásica, que es la base de la Estadística Matemática actual. Durante estos años, y a partir de ellos, se genera una notable producción estadística. Por un lado, se ha continuado trabajando en campos ya abiertos, como es el caso de la teoría de distribuciones, del muestreo, del análisis de la regresión, etc. Por otro lado, se iniciaron nuevos campos tales como el análisis secuencial, la teoría de la decisión, las técnicas multivariantes, el análisis de las series temporales, los métodos no para métricos, la estadística bayesiana, las técnicas para el control de calidad o el análisis exploratorio.

Los trabajos realizados por Neyman, junto a la Teoría de Juegos y la Teoría de la Utilidad legadas por J. Von Neuman (1903-1957), le sirvieron de punto de partida a A. Wald (1902-1950) para sentar las bases de la Teoría de la Decisión. Este nuevo enfoque, a cuya expansión contribuyeron los trabajos de A. Wald recogidos en *Statistical Decision Functions*⁶ en 1950, ha tenido una gran repercusión en los últimos cuarenta años. Casi todos los problemas de la inferencia estadística, y otros nuevos, se han reformulados en términos de esta teoría. Posteriormente, Savage extiende los resultados de Wald al introducir las probabilidades subjetivas y las funciones de utilidad, sentando las bases de la moderna Estadística Bayesiana, cuyas raíces históricas se encuentra en el teorema de Bayes de ahí su denominación.

En la Estadística Bayesiana, que en los últimos cincuenta años ha experimentado un enorme desarrollo, han incidido los planteamientos matemáticos y filosóficos contenidos en los trabajos, sobre Teoría de Juegos, de Borel, Von Neuman y Morgenstern, y sobre Probabilidad subjetiva de Ramsey y De Finetti. Más tarde, este nuevo enfoque ha sido desarrollado por estadísticos de la talla de I. J. Good, H. Jeffreys, D. V. Lindley o L. Savage, entre otros. La versión bayesiana de la Estadística tiene, al igual que la clásica, un sentido básicamente inferencial, en el que los conceptos de verosimilitud y suficiencia siguen manteniendo una importancia considerable. La diferencia fundamental entre

⁶ DeGroot (1988) califica esta obra como el primer manual en presentar la base matemática de la Teoría de la Decisión escrita por la misma persona que la había desarrollado en su mayor parte.

ambos métodos reside en los elementos aportados por la realidad a las técnicas inferenciales.

En el campo de Técnicas de Muestreo, cabe resaltar a Neyman por un trabajo sobre el muestreo en poblaciones finitas y por ser el pionero en el estudio de la técnica del muestreo doble que utiliza una muestra preliminar para estimar el tamaño de los estratos. En el estudio de Muestreo Estratificado y Sistemático, nombres como Cochran y Madow merecen especial atención. Así como Hansen y Hurwitz, quienes en 1942, introdujeron el muestreo por conglomerados.

Wald estudió por primera vez el muestreo de tipo secuencial. Los tests de tipo secuencial creados por él son muy utilizados en Control de Calidad. En su obra publicada en 1950, introduce la Teoría de la Decisión, por lo cual es considerado como uno de los pioneros de la misma. Su colaboración tuvo una fuerte incidencia en la teoría estadística de postguerra, en Teoría de Juegos y Teoría de la Decisión

En el campo del Análisis Multivariante, a partir de las aportaciones de Samuel Stanley Wilks (1906-1964) -verificación de hipótesis multivariantes-, Hotelling -análisis de componentes principales y de correlaciones canónicas-, Fisher -discriminación y clasificación- y de Mahalanobis -distancias generalizadas-, se han desarrollado técnicas de creciente complejidad, de forma que la aplicación de algunas de ellas -ciertos procedimientos de clustering o de escalado multivariante- no ha sido posible hasta disponer de medios informáticos.

El análisis de Series Temporales, esencial en Economía Cuantitativa, ha experimentado un gran desarrollo a partir de los trabajos de Yule en 1930 -procesos autorregresivos-. Una obra de gran repercusión en este campo ha sido *Time Series Analysis, Forecasting and Control* de Box and Jenkins publicada en 1970. En este campo, otras aportaciones importantes son la del análisis espectral, los modelos de interacción y filtrado, y la modelización multivariante, realizadas por autores como Harvey, Granger o Newbold, entre otros. Pero, sobre todo, es destacado por algunos autores, entre ellos Baxter (1990), que en algunas ramas de la Estadística el futuro vendrá condicionado por el papel predominante de los ordenadores y de la informática, tal es el caso de la modelización paramétrica, que podrá utilizar procedimientos más complejos, flexibles y sofisticados, como consecuencia de los avances en computación.

Por otra parte, en los últimos años, después de un largo periodo de axiomatización y matematización, los investigadores estadísticos han vuelto su mirada hacia la base de la Estadística, los datos. La publicación en 1977 de la obra *Exploratory Data Analysis* de J. W. Tukey supone el reconocimiento de la importancia de los datos dentro del método estadístico. En ella se incorporan nuevas técnicas gráficas y analíticas para el estudio previo de los datos. También en este ámbito, los métodos basados en simulación con ordenador parece que serán la vía para establecer una conexión entre la Estadística de base descriptiva y la de base probabilística.

Se observa, pues, como este siglo se puede calificar como un siglo capital en la historia de la Estadística, debido fundamentalmente a la retroalimentación de la misma

con otras ciencias y con la realidad. La Estadística es aplicada a la investigación de las diversas parcelas de la realidad y esta realidad ha supuesto una generación de nuevas ideas y perfeccionamientos de los métodos estadísticos. Por ello, parece razonable pensar que el futuro de esta disciplina vendrá condicionado, como ha sucedido con su evolución hasta el presente, por la necesidad de resolver los problemas prácticos que se vayan planteando.

Si difícil era efectuar una revisión de las principales aportaciones de la estadística a lo largo del presente siglo, tarea ardua es la de pronosticar su evolución en un futuro. A pesar de ello se pueden encontrar autores que expresan algunas consideraciones al respecto, entre ellos cabe citar a Peña (1986) quien enumera algunas metodologías que se desarrollaran en el futuro tales como:

- Técnicas exploratorias multivariantes para sintetizar la información de los grandes bancos de datos de acceso generalizado.
- Métodos más generales y flexibles de estimación.
- Métodos de evaluación de la incertidumbre, sobre todo en lo que se refiere al modelo de elección.
- Predicción de series multivariantes.
- Metodologías para meta-análisis que permitan combinar información con distinto grado de precisión.

4. METODOLOGÍA DE LA ESTADÍSTICA

Aunque entre los científicos se observan discrepancias respecto a la forma en que la Ciencia Estadística se concibe, no existe ninguna divergencia en considerar a la Estadística como ciencia instrumental cuyo campo de aplicación incluye entre otras a las ciencias sociales. Así Kendall y Stuart (1977), definen la Estadística como "*Una rama del método científico que se ocupa de los datos obtenidos contando o midiendo las propiedades de los fenómenos naturales*". Para Kendall (1968), la Estadística es el método científico por excelencia.

Atendiendo al carácter instrumental de la Estadística, ésta proporciona herramientas para la toma de decisiones cuando prevalecen condiciones de incertidumbre en cualquier campo del saber. Pero este carácter instrumental no implica que la Estadística no sea una ciencia formal.

A continuación nos ocupamos del método de investigación, de adquisición de conocimientos, en la Estadística. El carácter dual de esta disciplina, de ciencia formal y de ciencia aplicada, junto a su carácter instrumental, confieren a su metodología unas características especiales. Según algunos autores, las corrientes metodológicas principales que han llegado a nuestros días son dos: el método inductivo, asociado a las ciencias empíricas, y el deductivo, asociado a las formales. Algunas ciencias particulares participan de ambos métodos, como es el caso de la Estadística, la cual presenta un doble perfil: De ciencia aplicada y ciencia formal. En este sentido, el método estadístico se ha descrito como un proceso circular que implica una interacción cerrada y continua

entre intuición y observación así Box (1976) señala que el avance del conocimiento en Estadística surge de un bucle que se retroalimenta, y cuyos extremos son la Teoría y la Práctica. Ello supone que la Estadística incorpore elementos del razonamiento deductivo en una de las fases del ciclo, construcción de teorías y modelos, y en otra de las fases utilice elementos del método inductivo -formulación de hipótesis que son consecuencia de la observación de la realidad objeto de estudio-. Asimismo, destaca la opinión de R. A. Fisher, uno de los más grandes estadísticos de este siglo, cuando afirma que *" el análisis estadístico de un conjunto de datos es semejante en su lógica a la alternancia de métodos inductivos y deductivos típica de todas las ciencias. Una hipótesis es concebida y definida con toda la exactitud necesaria; sus consecuencias lógicas se derivan mediante un argumento deductivo; estas consecuencias se comparan con las observaciones disponibles; si éstas están completamente de acuerdo con las deducciones, la hipótesis es aceptada al menos hasta que se disponga de observaciones nuevas y más rigurosas"*.

Así, pues, en referencia al método estadístico, como proceso sistemático y organizado de la investigación científica, muchos son los autores y docentes que opinan que emplea un razonamiento deductivo-inductivo interactivo. El esquema descrito debe ser lo suficientemente flexible como para permitirnos el reconocimiento de los errores inherentes a la teoría inicial, de modo que no corra el riesgo de idealizar el modelo propuesto.

El carácter dual de la Estadística, reflejado en la dualidad de su método, ha llevado a que algunos autores la hayan considerado como una ciencia matemática, mientras que para otros se trata de una herramienta en la investigación científica (Bradley, 1982) y Kish (1978) consideran que la Estadística sólo existe en la interacción con los datos empíricos, al tiempo que no tiene un campo propio donde encontrar esos datos. Posiblemente este tipo de afirmaciones, un tanto tajantes, se deba, en gran parte, a que la distribución de investigadores en esta disciplina sea bimodal: existe una gran concentración de investigadores muy especializados en aspectos teóricos y otro gran número de ellos dedicados a tareas aplicadas, con un vacío de comunicación entre unos y otros.

En lo que se refiere a la Estadística como instrumento, su papel central en el campo científico (producción de datos, descripción, diseño de experimentos y muestras, estimación, verificación, predicción, análisis de series, teoría de la decisión, etc.) es cada vez más importante, lo que ha dado lugar al nacimiento de ramas especializadas (Biometría, Sociometría, Econometría,) de forma que sus límites se difuminan, lo que, como señala Neter (1986), es una característica positiva si se tiene en cuenta que, en gran medida, los desarrollos de la Estadística han tenido lugar como consecuencia de la interacción con otras ciencias, lo que ha proporcionado el impulso necesario para el desarrollo de nuevas técnicas. En este aspecto, nuestra disciplina es una valiosísima ayuda en la contrastación y validación de teorías, incluso para fijar los límites de la misma, pero no puede utilizarse para la formulación de las teorías.

El punto de partida de una investigación estadística es, en sentido amplio, la consideración de una situación real en ambiente de incertidumbre. De su observación se extraen los aspectos fundamentales, los cuales se formalizan para construir un modelo general que recoja la naturaleza del objeto de investigación, pero dado el contexto de incertidumbre, el anterior modelo hay que mantenerlo en términos probabilísticos. Posteriormente, la contrastación del modelo con la realidad conduce a aceptar, modificar, reelaborar o sustituir el modelo. Respecto al modelo, para su construcción se parte de unos principios básicos, de algún modo extraídos de la realidad a partir de la experiencia, y sobre la base de un proceso deductivo se elabora un entramado lógico formal que ayuda a construir el modelo. El principal problema que plantea ese camino hacia la esquematización de lo real es el paso de la experiencia al sistema axiomático o de principios básicos. Los dos enfoques más importantes, sobre esta cuestión, son: el de los conductivistas y el de los falsacionistas. Para los primeros, el tránsito se lleva a cabo mediante la generalización de experiencias particulares. Y para los segundos debe ser mediante la falsación o contrastación de las conclusiones extraídas de ciertos principios intuitivos, como se recoja la conexión entre axiomas y realidad.

Por regla general todo proceso de análisis estadístico puede diferenciarse tres fases perfectamente estructuradas que un investigador ha de seguir. En la primera, se recogerá la información a través de los datos observados y que proporcionará información sobre el objeto de estudio. En la segunda fase, se propondrá el modelo de probabilidad en función de la información que se haya desprendido de las observaciones y en la tercera y última fase se efectuará un proceso de validación del modelo.

Los objetivos y las técnicas más utilizadas, en cada fase son:

- Fase Descriptiva.

En esta primera fase se recoge toda la información disponible a través de la observación de los datos. El conjunto de estas observaciones es lo que denominamos muestra, y sobre la que posteriormente se realizará el análisis. La elección de las observaciones que constituirán la muestra va a ser muy importante. El diseño muestral resultante, debe garantizar que la muestra sea representativa de la población, así como evitar errores o anomalías en los datos seleccionados.

Dada la importancia de la representatividad de la muestra, la Teoría de Diseño de Muestras ha desarrollado distintas técnicas para adecuarlas al problema a tratar. Habitualmente en la práctica se utiliza una técnica de fácil aplicación, como es el procedimiento de muestreo aleatorio simple. En él se considera que todas las componentes de la muestra tienen la misma probabilidad de ser escogidas y además se garantiza la existencia de independencia y equidistribución entre las variables observadas.

No obstante existen problemas en los que para poder abordarlos, se exige que el proceso de muestreo sea otro al anteriormente mencionado. Es el objetivo del estudio el que debe determinar el proceso de selección muestral para la obtención de la muestra.

De este proceso depende en gran parte que las conclusiones sobre el análisis de las observaciones sean coherentes con la realidad que se estudia. Debe tenerse presente que el diseño muestral y el análisis de los datos están estrechamente relacionados. Un buen diseño debe ir acompañado del consecuente análisis de los datos muestrales pues de lo contrario las conclusiones serán poco fiables. Si se realiza un análisis exhaustivo sobre datos muestrales no representativos, de nada servirán los resultados que se obtengan.

Un problema al que puede enfrentarse el investigador es cuando no puede decidir el diseño muestral, no puede participar en la obtención de la muestra ni garantizar su aleatoriedad. En tal caso, el estadístico ha de aceptar los datos tal y como están. Generalmente deberá admitir que los datos constituyen una muestra aleatoria y tratar de localizar las observaciones aberrantes o "outliers" para eliminar distorsiones sobre la realidad.

Seleccionada la muestra y considerados todos los posibles problemas que hemos mencionado, se procede al análisis descriptivo de los datos escogidos. Esta información recogida permitirá al estudioso continuar el análisis hasta modelizar la realidad o bien otras veces puede ser esta fase la única objeto de estudio.

- *Fase de Análisis.*

Una vez recogida la información en la etapa anterior, se intenta especificar un modelo que explique la realidad de forma adecuada. Generalmente cuando se propone un modelo, se pretende que cumpla dos condiciones: describir la realidad y evitar una excesiva complejidad en su composición estructural. A veces resultará complicado que se cumplan ambas condiciones ya que a menudo, elevar la fidelidad implica introducir hipótesis que complican el análisis. El estadístico deberá flexibilizar los modelos para poder tratarlos matemáticamente con mayor facilidad. A su vez exigirá que los resultados sean coherentes. Con ayuda de la Inferencia Estadística, el investigador aplica técnicas que le permiten realizar conclusiones lógicas con los datos observados.

Una primera forma de abordar el problema que se estudia consiste en determinar la distribución de probabilidad del modelo generador de datos que se supone perfectamente conocida a excepción del valor de algún parámetro. Con la información disponible sobre las observaciones muestrales, el investigador tratará de obtener el valor de esos parámetros desconocidos : necesarios para especificar completamente la distribución poblacional. De este proceso a través del cual se estiman los valores de los parámetros desconocidos, se encarga la Inferencia Paramétrica.

Es posible que el investigador no pueda establecer ninguna hipótesis I sobre el modelo generador. En tal caso necesitará realizar inferencias sobre I toda la distribución de probabilidad muestral. De las técnicas estadísticas necesarias para abordar el estudio según este planteamiento se encarga la Inferencia no Paramétrica. La Inferencia Paramétrica aborda la estimación de los parámetros desconocidos del modelo, puntualmente mediante la búsqueda de propiedades deseables y a través de intervalos

de confianza, estableciendo garantías de que el verdadero valor del parámetro oscilará entre dos posibles cantidades (mínima y máxima).

A la Inferencia no Paramétrica le corresponde entre otros aspectos, plantear la distribución de probabilidad más adecuada a los datos disponibles, establecer relaciones entre las observaciones o estudiar características de interés.

Un investigador puede tomar dos caminos diferentes cuando pretende proponer un modelo para la estructura poblacional objeto de estudio. Estos dos caminos se corresponden con dos concepciones diferentes de la Estadística. La clásica, en la que sólo se utiliza la información proporcionada por los datos observados y la bayesiana, donde se incorpora, además, la opinión del investigador sobre el problema. El componente subjetivo se considera, por tanto, una fuente de información en la Inferencia Bayesiana. En un análisis bayesiano, la información subjetiva del investigador se introduce a través de un modelo probabilístico que refleje sus opiniones.

Al incorporar tanto información subjetiva como objetiva, los modelos bayesianos suelen ser más complejos de operar. Como contrapartida es posible que al disponer de mayor información los modelos obtenidos se aproximen más a la realidad. Pese a todo existe evidencia suficiente de que utilizando las dos concepciones, un investigador puede llegar a las mismas conclusiones respecto al problema abordado. De todas formas, la uniformidad de resultados no es lo más habitual ya que tanto la concepción metodológica, las técnicas utilizadas en cada caso, como la información disponible serán diferentes. No obstante, si la información fuera la misma, los resultados obtenidos por ambas concepciones deberían ser coincidentes.

- Fase de Validación.

Una vez obtenido el modelo que explique la realidad objeto de estudio, el investigador ha de comprobar si dicho modelo proporciona resultados coherentes, es decir si es teóricamente aceptable y estadísticamente significativo. La Estadística proporciona las técnicas necesarias para llevar a cabo esta fase. Por ejemplo el contraste de Hipótesis Paramétrico y No Paramétrico y la Teoría de la Decisión se utilizan para este fin. Es preciso que todas las hipótesis consideradas en el modelo propuesto, se contrasten, así como la adecuación de las observaciones al mismo. Si se comprueba que no se ajusta a la realidad se tendrá que reiniciar todo el proceso.

Generalmente el investigador suele utilizar las técnicas de simulación para comprobar si el modelo es o no aceptable. A través de éstas genera valores con el modelo y comprueba si se aproximan a los observados. Cuando las discrepancias no son excesivas, se acepta el modelo como válido. Ahora bien, el estadístico deberá ser cuidadoso en dicha aceptación ya que podría ocurrir que el modelo sólo fuera válido para los datos muestrales.

Realizadas todas las contrastaciones, la aceptación del modelo permitirá dar una explicación de la realidad que se ha estudiado y que va a servir por ejemplo para dar

predicciones. Concretando más, la investigación estadística se desarrolla utilizando el ciclo deductivo-inductivo, que siguiendo a Peña Sánchez de Rivera (1991) consta de las siguientes etapas:

- Planteamiento del problema: El primer paso de la investigación es definir los objetivos de estudio y relacionar este objetivo con los valores numéricos de variables observables.
- Construcción del modelo estadístico: El tipo de modelo estadístico más frecuente consiste en una descomposición de los valores de la variable de interés en una parte sistemática más una parte aleatoria. Un modelo estadístico es una aproximación operativa de la realidad, que tiene en cuenta explícitamente las variables observables presumiblemente más importantes, y engloba en la parte aleatoria los efectos del resto. Podemos distinguir entre modelos extrapolativos, que explican la variabilidad de los datos tomando como información únicamente sus valores pasados, y modelos explicativos, que tienen en cuenta el efecto de otras variables.
- Recogida de la información muestral: Una vez construido un modelo del problema tendrá que medir los valores de las variables de interés. Esta recogida de información puede hacerse mediante muestreo o con un diseño de experimentos. El primero consiste en observar pasivamente una muestra de las variables y anotar sus valores; se utiliza en modelos extrapolativos. El segundo método, el diseño de experimentos, consiste en fijar los valores de ciertas variables y observar la respuesta de otras; debe emplearse cuando se desee construir un modelo explicativo.
- Depuración de la muestra: Antes de utilizar los datos muestrales conviene aplicar técnicas estadísticas simples para identificar errores, datos anómalos, etc.
- Estimación de los parámetros: Los modelos estadísticos dependen de ciertas constantes desconocidas que llamaremos parámetros. La fase de estimación consiste en utilizar la información disponible para decidir respecto a un valor concreto de estos parámetros así como cuantificar el posible error en la estimación.
- Contrastes de simplificación: El objetivo de esta fase es conseguir un modelo tan simple como sea posible.
- Crítica y diagnosis del modelo: Los resultados de las dos etapas anteriores se obtienen suponiendo que el modelo es correcto. Esta fase investiga la compatibilidad entre la información empírica y el modelo. Si después de esta fase aceptamos el modelo como correcto, lo utilizaremos para tomar decisiones o realizar previsiones de la variable. En caso contrario volveremos a reformular el modelo estadístico, repitiendo el proceso hasta conseguir un modelo correcto.

Finalmente, de manera amplia y general, se puede decir que la característica esencial de la investigación estadística desde el punto de vista metodológico es su participación del método axiomático y de las connotaciones empíricas que caracterizan las ciencias, que como toda ciencia factual, trata de reflejar aspectos de la realidad circundante. En concreto, la Estadística Económico-Empresarial usa del método deductivo para crear teorías y modelos generales, y del método inductivo para estudiar la realidad o la parte de la realidad que constituye su objeto, es decir, los fenómenos de azar del mundo empresarial. Este proceso se constituye en un procedimiento de retroalimentación para la aceptación o el rechazo de los modelos planteados.

Referencias:

BAXTER, L. A. (1990): "Futures of Statistics", *The American Statistician*, Vol. 44, n° 2,

BOX, G. E. P. (1976): "Science and Statistics", *Journal of the American Statistical Association*, vol. 71, n° 356,

BOX, G.E.P. (1979): "Some Problems of Statistics and Everyday Life", *Journal of the American Statistical Association*, n° 365.

BOX, G.E.P. (1990): "Comment on: 'Applications in Business and Economic Statistics: Some Personal Views'", Roberts, H.V. *Statistical Science*, Vol 5-4.

BRADLEY, R. A. (1982): "The Future of Statistics as a Discipline", *Journal of the American Statistical Association*, Vol. 77, n° 377,

CRAMER, H. (1968): *Métodos Matemáticos de Estadística*. Aguilar. Madrid

CRAMER, H. (1981): "Mathematical Probability and Statistical Inference. Some Personal Recollections from an Important Phase of Scientific Development", *International Statistical Review*, Vol. 49,

DEGROOT, M.H. (1970): *Optimal Statistical Decision*, McGraw-Hill, New York.

DEGROOT, M.H. (1988): *Probabilidad y Estadística*, Addison-Wesley, México.

ESCUDE R VALLÉS, R. (1987): *Métodos Estadísticos Aplicados a la Economía*, Ariel, Barcelona.

FERNANDEZ ABASCAL, H. ; GUIJARRO, M.; ROJO, J. L. Y SANZ, J. A. (1994): *Cálculo de Probabilidades y Estadística*, Ariel, Barcelona

FERNANDEZ DE TROCONIZ, A. (1980): *Introducción a la Teoría de las Probabilidades. Estadística Clásica y Bayesiana*. Editado por el Autor. Bilbao

FERNANDEZ DE TROCONIZ, A.. (1982): *Probabilidades y Métodos Estadísticos*. Editado por el Autor, Bilbao.

GUTIÉRREZ CABRÍA, S. (1968): "Consideraciones en torno al Concepto de Probabilidad e Incertidumbre", *Estadística Española*, n° 41

GUTIÉRREZ CABRÍA, S. (1982a): "Inferencia Estadística e Inducción", *Estadística Española*, n° 94,

GUTIÉRREZ CABRÍA, S. (1982b): "Origen y Desarrollo de la Estadística en I los siglos XVII y XVIII", *Estadística Española*, n° 97,

GUTIÉRREZ CABRÍA, S. (1983a): "Desarrollo de la Inferencia Estadística .desde sus Comienzos hasta Principios de este Siglo", *Estadística Española*, P n° 98,

HACKING, I. (1995): *El Surgimiento de la Probabilidad*, Gedisa, Barcelona.

HEILBRONER, R.L. (1988): *La formación de la sociedad económica*, Fondo de Cultura Económica, México.

KENDALL, M. G. (1968): "Francis Ysidro Edgeworth, 1845-1926", *Studies in the History of Probability*, Charles Griffin, Londres

- KENDALL, M. G. (1968): "Sobre el futuro de la Estadística. Una segunda visión" *Estadística Española*, nº 41,
- KENDALL, M.G. y STUART, A. (1977) *The Advanced Theory of Statistics. Vol I: Distribution Theory*. London and High Wycombe: Charles Griffin & Co. Ltd.
- KENDALL, M. G. (1979): "Estadística. Historia del Método Estadístico", en SILLS, d. (dir.) *Enciclopedia Internacional de las Ciencias Sociales*, Aguilar, Madrid
- KISH, L. (1978): "Chance, Statistics and Statisticians", *Journal of the American Statistical Association*, nº 361
- KATOUZIAN, H. (1982): *Ideología y método en economía*, H. Blume Ediciones, Madrid.
- MOOD, A. y GAYBILL, F.A. (1978): *Introducción a la Teoría de la Estadística*. Aguilar. Madrid
- NETER, J. (1990): "Comment on: Applications in Business and Economic Statistics: Some Personal Views", *Statistical Science*, vol. 5, nº 4,
- NEWBOLD, P. (1998): *Estadística para los Negocios y la Economía*, Prentice Hall, Madrid.
- PEÑA SÁNCHEZ DE RIVERA, D. (1986): *Estadística, Modelos y Métodos*, Alianza, Madrid.
- PEÑA SÁNCHEZ DE RIVERA, D. (1992) "Reflexiones sobre la Enseñanza Experimental de la Estadística", *Estadística Española*, nº 32.
- ROBBINS, L(1935):*An Essay on the Nature and Significance of Economic Science*. Macmillan. Londres (fragmento compilado en Kahn F., Hollis, M. *Phylosophy and Economic Theory*. Oxford UP. 1979. -traducida al castellano en 1986 por FCE-)
- THUROW, L. (1988): *Corrientes peligrosas. El estado de la ciencia económica*, Fondo de Cultura Económica, México.
-