

DESARROLLO HISTÓRICO DE LA ESTADÍSTICA

Extractado por Jorge Galbiati Riesco

Las fuentes de la Estadística la constituyen los censos y recuentos, los juegos de azar, la inferencia inductiva basada en datos empíricos, y el tratamiento de los errores en las mediciones.

Se sabe que Cesar Augusto decretó que todo el imperio fuera sometido al pago de impuestos, para lo cual previamente debería conducirse un censo de las personas. Mil años después, Guillermo el Conquistador ordenó que se hiciera un registro de todos los bienes que hubieran en Inglaterra, para fines tributarios y militares, el llamado "Domesday Book". Una aplicación de la probabilidad empírica a los seguros de buques se encuentra en Flandes, en el siglo XIV.

La teoría de la probabilidad es una disciplina matemática que fundamenta la Estadística como una lógica y una metodología para la medición y el estudio de la incertidumbre, en la planeación e interpretación de la observación y la experimentación.

Los inicios de la probabilidad, como teoría matemática, puede rastrearse en la correspondencia entre Fermat y Pascal, en la década de 1650. Pierre de Fermat, matemático francés, nació en 1601; Blaise Pascal, matemático, físico y filósofo, también francés, nació en Clermont-Ferrand en 1623.

También hay antecedentes de los orígenes de la teoría de la probabilidad en un corto artículo escrito por Christian Huygens en 1657. Fue éste un físico, geómetra y astrónomo holandés, nacido en La Haya en 1629. Previamente, Girolamo Cardano (1501-1576) y Galileo Galilei (1564-1642) habían hecho cálculos de probabilidades numéricas, de diversas combinaciones de dados.

Estos trabajos tempranos de Fermat, Pascal y Huygens no abordan problemas de estadística inferencial, ni van más allá de los juegos de azar, que eran sus intereses inmediatos.

Un comerciante inglés, John Graunt, publicó en 1662 un artículo titulado "Natural and Political Observations upon the Bills of Mortality", en el que presenta cálculos demográficos que evidencian el reciente descubrimiento de la regularidad de ciertas proporciones. Pasarían décadas antes que se tuviera conciencia de la existencia de variabilidad de todos los fenómenos. Por sus trabajos en demografía, que incorporan nociones de regularidad en el comportamiento de características de naturaleza aleatoria, John Graunt es considerado por algunos, como el iniciador de la Estadística. Fue socio fundador de la Royal Society.

Jacob Bernoulli, (o James, o Jacques Bernoulli), un matemático suizo nacido en 1654, es considerado el iniciador de la teoría de la probabilidad, que hasta entonces sólo se había ocupado de fenómenos experimentales con resultados equiprobables, motivados, aparte de los juegos de azar, por problemas de las ciencias sociales, intereses Financieros, seguros, meteorología y medicina. En su obra "Ars Conjectandi", introduce lo que hoy se conoce como la primera ley de los grandes números. Es éste un principio fundamental, que básicamente establece que, bajo ciertas condiciones, un promedio muestral se aproxima al promedio de la población de donde se obtuvo la muestra, si el tamaño de ésta es grande.

Entre los siglos XVIII y XIX, la Estadística experimentó lo que puede ser descrito como un desarrollo horizontal y vertical simultáneo. Horizontal, en el sentido que se propagó a través de diversas disciplinas, desde la astronomía y la geodesia, la psicología, la biología, hasta las ciencias sociales, sufriendo diversas transformaciones en el proceso. Vertical, en el sentido de profundizar en el conocimiento del rol de la probabilidad, siendo desplazada la analogía de los juegos de azar, por modelos probabilísticos apropiados para efectuar medidas bajo incertidumbre. De este modo se llega a los inicios de la inferencia estadística, cuyo dominio de aplicación se extiende gradualmente, desde fines de este período.

Un matemático francés nacido en 1667, que vivía en Londres, refugiado de la persecución religiosa, Abraham De Moivre, publicó tres obras de contenidos sobre el tema de la probabilidad, entre 1718 y 1730. Contribuyó efectuando estudios sobre la ley de probabilidad binomial, y formuló una aproximación para muestras grandes, que es considerada por estadísticos de este siglo, como Karl Pearson, como la primera formulación de la ley de probabilidad normal. Pearson encontró la publicación de De Moivre en que presentaba la ley normal. Pero el autor no había descubierto su aplicación a la

descripción del error en observaciones experimentales. Fueron Laplace y Gauss quienes lo hicieron, independientemente, un siglo después.

En 1761 muere un inglés, Thomas Bayes, ordenado ministro, pero con interés en la matemática, sin haber publicado un sólo trabajo matemático durante su vida. Su obra "Ensayo sobre la Resolución de un Problema en la Doctrina del Azar", publicada póstumamente en 1764, fue ignorada por sus contemporáneos, y parece haber tenido poca influencia sobre el desarrollo temprano de la Estadística. Irónicamente, sus contenidos, estudios de la inversión de la probabilidad, sirvieron, dos siglos después, para grabar su nombre en toda una corriente estadística, la moderna inferencia bayesiana.

La inferencia bayesiana permite asignar probabilidades a fenómenos que no son de naturaleza aleatoria, pero cuyos resultados no son conocidos, lo que no es posible bajo el punto de vista opuesto, el de los frequentistas, que sólo permiten asignar probabilidades cuando es posible que éstas puedan ser apoyadas por la experimentación. El Teorema de Bayes logra afinar la asignación de estas probabilidades, a medida que se adquiere conocimiento de la población bajo estudio, a través de la obtención de observaciones.

En 1763 un inglés, Arthur Young, heredó un fundo, y en él comenzó a experimentar para descubrir el método agrícola más rentable. Desarrolló un gran número de experimentos, publicando sus resultados en un libro llamado "Un Curso de Agricultura Experimental", en 1771. Las ideas que presenta sobre el Diseño de Experimentos, una importante disciplina de la Estadística actual, cuyas aplicaciones al campo industrial se encuentran hoy en pleno desarrollo, son sorprendentemente modernas.

El físico, matemático, y astrónomo francés Pierre-Simon Laplace, hizo contribuciones importantes a la aplicación de la probabilidad a la inferencia estadística, contenidas fundamentalmente en dos obras "Memoria sobre la Probabilidad de las Causas de Eventos", de 1774, y "Memoria sobre Probabilidades", de 1781. Se preocupó de la inversión de la probabilidad, como Bayes, pero sin conocer los resultados obtenidos por este último, llegando a formular un caso particular del teorema de Bayes, con la adopción tácita de probabilidades a priori iguales, con posterioridad a la publicación póstuma de la obra de Bayes. Contribuyó en muchos temas estadísticos, entre ellos en la obtención de una "curva de errores", llegando a la formulación de la ley de probabilidad normal.

En 1805, el matemático francés Adrien Marie Legendre dio a conocer un sistema para determinar las órbitas de los cometas, que involucra una descripción del método de los mínimos cuadrados, tan utilizado en la Estadística de hoy. Es un método de estimación de parámetros, que básicamente consiste en asignarles los valores que minimizan la suma de los cuadrados de las diferencias entre las observaciones y los valores teóricos entregados por el modelo propuesto. El método de los mínimos cuadrados fue el tema dominante de los estadísticos del siglo XIX.

Karl Friederich Gauss, un matemático nacido en Alemania, en 1777, también interesado en el estudio de las órbitas de los planetas, contribuyó al método de los mínimos cuadrados, desembocando, independientemente de Laplace, en la ley de probabilidad normal, o curva de Gauss, como descripción probabilística del error. Pero Gauss encontró su asociación con el método de mínimos cuadrados. Hacia 1830, Adolphe Quetelet, astrónomo, meteorólogo, estadístico y sociólogo, hizo los primeros intentos de aplicar la probabilidad a la medición de la incertidumbre en las ciencias sociales. Su avance hacia el análisis estadístico de datos sociales fue el hecho de introducir el concepto del hombre promedio, motivado por sus investigaciones de datos demográficos, observando múltiples relaciones entre características contenidas en datos de poblaciones humanas. Su contribución más perdurable fue el hecho de ajustar distribuciones de probabilidad a datos empíricos.

Adolphe Quetelet era un belga, nacido en Gantes, en 1796, interesado en las bellas artes. Se dedicó a la pintura, escribió poesías, e incluso escribió una ópera. Pero después desarrolló una inclinación a las matemáticas, que lo llevó a interesarse por el estudio de la teoría de la probabilidad y su aplicación a los fenómenos sociales. Es así como contribuyó a impulsar la realización del primer censo nacional en Bélgica y Holanda, e hizo esfuerzos por que se uniformizaran los métodos y la tecnología utilizada en la recolección y presentación de datos, en Europa. Tuvo un liderazgo importante en la creación de organizaciones ligadas a la Estadística, como la Statistical Society of London, ahora Royal Statistical Society.

A Quetelet se le ha llamado el "padre de la Estadística moderna", por una publicación suya, de 1835, en que observa la extraordinaria regularidad con que se reproducían ciertos fenómenos sociales, como crímenes o suicidios, y argumenta que esas regularidades sólo pueden ser encontradas mediante el uso de técnicas estadísticas, las que incluso pueden llevar a conocer sus causas. Quetelet pensaba que casi todos los

fenómenos pueden ser representados probabilísticamente mediante la ley normal, siempre que el número de casos estudiados fuese suficientemente grande.

Simeón Denis Poisson, físico matemático nacido en Francia, en 1781, publicó un gran tratado de probabilidad en 1837. Contiene el germen de dos elementos asociados al nombre de Poisson: La ley de probabilidad conocida como distribución de Poisson, y la generalización de la ley de los grandes números de Bernoulli.

Numerosos investigadores, provenientes de las más diversas disciplinas, hicieron contribuciones a la Estadística, durante la segunda mitad del siglo XIX, construyendo de a poco una disciplina que se iría perfilando cada vez más como una ciencia independiente.

Wilhelm Lexis, economista alemán, contribuyó a la estadística social, estudiando datos presentados como series a través del tiempo, hacia 1880. Se inicia un tema importante dentro de la Estadística, el de las series de tiempo, muy utilizadas hoy en día, en particular, en sus aplicaciones a la Economía.

John Arbuthnot, médico de la reina Ana de Inglaterra, es más conocido como estadístico, por sus estudios sobre las proporciones de los sexos en los nacimientos.

Henry Buckle, inglés, precursor de la moderna ciencia histórica, utilizó métodos estadísticos para ayudar a hacer de la historia una ciencia.

Gustav Fechner, alemán con estudios Incompletos en medicina, incursionó *en* las técnicas de la experimentación, para describir la relación entre estímulos y sensación, derivando la Estadística hacia el campo de la psicología experimental. Fechner introdujo la medición en la psicología. Aparentemente creó el término psicofísica para describir la psicología experimental, que practicó como disciplina formal, a mediados del siglo XIX.

Un psicólogo, Hermann Ebbinghaus, siguió la línea de Fechner, compartiendo dos ideas cruciales: Que el estudio cuantitativo era la única manera de expresar con precisión las vagas nociones que la psicología manejaba antes, y el convencimiento que el apoyo de un cuidadoso diseño experimental es fundamental para la experimentación. Ebbinghaus aplicó estas ideas al estudio de la memoria.

A partir de 1880, tres hombres, Francis Galton, Francis Edgeworth y Karl Pearson, crean una revolución en la Estadística, proporcionando una metodología empírica que sustituye a la experimentación controlada, en disciplinas donde la experimentación no es posible de aplicar. Esta metodología empírica ya había sido utilizada en la psicología. Lo hicieron separadamente Galton en la Antropología, Edgeworth en la Economía y Pearson en la filosofía de la ciencia.

Francis Galton, nacido en 1822, investigó el carácter hereditario de la genialidad, utilizando curvas normales inversas, que llamó "ojivas", término que tomó de la arquitectura, y que aun se utiliza. Fue pionero en el tema de la regresión lineal simple, o reversión, como él la llamó, técnica para obtener una expresión que relaciona en forma lineal, dos características. También se preocupó de la estimación de las componentes de varianza, o partes de la variabilidad de un fenómeno observado, atribuibles a causas identificables. El concepto estadístico por el que es más conocido es el de la correlación, aún cuando su interpretación limitada a un coeficiente que mide el grado de asociación entre el comportamiento de dos variables, tan utilizado en el presente, jugó un papel poco importante en el trabajo de Galton.

Francis Galton también utilizó la ley de probabilidad normal, en su versión bivariada, para describir el comportamiento probabilístico de los errores de dos características que varían en forma conjunta. Esta ya había sido conocida por otros investigadores, desde principios del siglo XVIII: Por el estadounidense Robert Adrain en 1808, por Laplace en 1811, y por Gauss en 1823, entre otros. Pero ellos no conocieron el coeficiente de correlación, presente en la normal bivariada, y que cuantifica el grado de asociación entre las dos características. La ley normal bivariada daría origen a la ley normal multivariada. Esta es fundamental en la rama de la estadística denominada análisis multivariante, que se preocupa del estudio de observaciones con múltiples variables.

Francis Ysidro Edgeworth se educó en literatura clásica, luego estudió derecho comercial. Pero después se interesó en aplicar los métodos estadísticos previamente aplicados en astronomía y geodesia, a la economía y sociología. Contribuyó al desarrollo de la regresión y la correlación. Edgeworth también trabajó en otra área fundamental de la Estadística, que estudia las aproximaciones que se obtienen cuando los conjuntos de datos crecen ilimitadamente. Aporta la aproximación de Edgeworth, cuya aplicación se ha intensificado en la actualidad. Desarrolló una versión del teorema del límite central, una

herramienta muy utilizada, que establece, en líneas generales, que bajo ciertas condiciones, un promedio muestral sigue aproximadamente la ley probabilística normal, sin importar que comportamiento probabilístico tiene la población de donde provienen las observaciones, si el número de observaciones es grande.

Karl Pearson mostró interés en distribuciones probabilísticas asimétricas, en contraposición con las distribuciones normales, simétricas. Llegó de esta manera a introducir una familia de distribuciones probabilísticas, hoy conocida como Gama, que ya había sido descubierta independientemente por el estadounidense Erastus De Forest. Pearson mostró interés en los más diversos temas, además de la estadística, como la filosofía, la religión, la historia, entre otros. Su "Gramática de las Ciencias", de 1892, ilustra su convicción de que la estadística analítica yace en los fundamentos de todo el conocimiento.

Pearson, en su trabajo, dio más importancia a la cuantificación de la correlación entre dos variables, en la forma de un coeficiente, que la que le había dado Galton. El y otros investigadores desarrollaron varios coeficientes de correlación, para el estudio de diferentes problemas en genética, biología, y otras disciplinas. El más común y conocido de ellos, hoy en día, lleva su nombre. A Karl Pearson se debe, también, el estadístico ji-cuadrado, introducido en 1900. Este estadístico, es utilizado como medida de comparación entre dos tablas de frecuencia, y una de sus aplicaciones es el probar el ajuste de una ley probabilística a un conjunto de datos empíricos.

George Udny Yule, inglés con estudios de ingeniería y física, fue un colaborador de Pearson, que hizo algunos aportes a la obra de este último. Trabajó en correlación, y también en curvas asimétricas, como su predecesor. Colaboró en la publicación de Pearson, proporcionando un ejemplo de la aplicación de ajuste de una curva asimétrica a datos sobre distribución de pobreza en Inglaterra y Gales. Pero luego se movió en direcciones independientes. Relacionó la regresión con el método de los mínimos cuadrados, proporcionando un gran conjunto de algoritmos que habían desarrollado los astrónomos, para la solución de las ecuaciones normales, asociadas al cálculo de la regresión. Los trabajos publicados por Yule cubren hasta la primera década de este siglo.

La idea de representatividad, en Estadística, es decir, de seleccionar aleatoriamente algunas unidades para llevar a cabo un estudio sobre una población, es

antigua. En esta idea se fundamenta la técnica de muestreo. Sin embargo, durante mucho tiempo no fue aceptado, por la generalidad de los estadísticos. En 1895, fue presentada formalmente en una reunión del Instituto Internacional de Estadística, realizada en Berna, por el director de la Oficina Central de Estadística de Noruega, A. N. Kaier, bajo el nombre de método representativo, en contraposición a la investigación exhaustiva.

Despertó gran interés, pero finalmente fue rechazado. No hubo más informes a favor del método de muestreo sino hasta la reunión del Instituto Internacional de Estadística celebrada en Roma, en 1926. Tuvieron influencia favorable los trabajos sobre representatividad en estudios sociales y económicos, debidos a A. L. Bowley. A él se debe una aplicación de la teoría de inferencia a las encuestas por muestreo, hecha en 1906. Aplicó el teorema del límite central, en la versión de Edgeworth, basada en el teorema de Bayes, para evaluar la precisión de las estimaciones obtenidas con grandes muestras aleatorias de poblaciones grandes, finitas. Tippett publica la primera tabla de números aleatorios, en 1927, para la obtención de muestras al azar.

En 1934, el polaco Jerzy Neyman publicó en la Royal Statistical Society, de Londres, lo que puede ser considerado el primer trabajo científico sobre el muestreo de poblaciones finitas. Estableció, sin lugar a dudas, que la selección aleatoria es la base de una teoría científica que permite predecir la validez de las estimaciones muestrales. También dejó establecida toda una filosofía sobre la eficiencia de la estrategia muestral. Neyman y Egon Pearson, hijo de Karl Pearson, presentaron en 1936 una teoría sobre la forma de probar hipótesis estadísticas, en base a datos. Esta presentación promovió mucho interés, estimuló una considerable cantidad de investigación, y muchos de los resultados hasta hoy aun se usan. Ellos resolvieron dificultades fundamentales para la comprensión de las pruebas de hipótesis, introduciendo las nociones de hipótesis alternativa, y los dos tipos de error, el de rechazar una hipótesis que es verdadera, y el de no rechazar una hipótesis que es falsa. Surge un resultado fundamental, el Lema de Neyman-Pearson, y se crea una larga controversia con R. A. Fisher, que visualizaba la prueba de hipótesis como un procedimiento mediante el cual el investigador podía formarse una opinión sobre alguna característica de la población, o parámetro. Neyman y Pearson vieron la prueba de hipótesis como un medio para que el investigador tomara una decisión sobre un parámetro de la población.

Neyman introdujo, en 1934, la teoría de los intervalos de confianza. Es una forma de estimar un parámetro, contrapuesta a la estimación puntual, que determina un intervalo que contiene el parámetro, y un coeficiente de confianza, que representa la probabilidad que el intervalo efectivamente contenga al parámetro. Los intervalos de confianza y las pruebas de hipótesis son dos elementos de la inferencia estadística.

En las décadas de los años 30 y 40 se centra el desarrollo de la técnica del muestreo estratificado, que asume que existen segmentos distintos, o estratos, en la población, que pueden ser identificados previamente, los cuales se muestrean separadamente. Jerzy Neyman fue un arduo defensor del método de muestreo estratificado, y su trabajo en este tema, abrió nuevas áreas de investigación.

También en la década del 40 se desarrolla el método de muestreo por conglomerados, que consiste en muestrear grupos, en lugar de unidades, para luego censar estos grupos. En este periodo se establecen las condiciones bajo las cuales estos métodos, el muestreo estratificado y el muestreo por conglomerados, resultan más eficientes.

En 1908, el inglés William Gosset, quien fuera alumno de Pearson, publica un artículo "El Error Probable de una Media", bajo el seudónimo de Student. Este artículo constituye un paso importante en el sentido de cuantificar los resultados de la experimentación. No está claro cuando Gosset se interesó por la Estadística, pero trabajaba en la cervecería Guinness, en cuyo ámbito se encontraba con problemas relacionados con muestras pequeñas, para las cuales la teoría de muestras grandes, existente entonces, proporcionaba sólo una mala aproximación. Esto lo llevó a desarrollar la ley probabilística que hoy es conocida como t de Student, utilizada en lugar de la ley normal de Gauss, en problemas con muestras pequeñas. Y también lo llevó a desarrollar la prueba de hipótesis llamada hoy test de Student, para inferencias sobre medias poblacionales, basadas en muestras pequeñas.

Sin embargo, pasaron largos años antes que el test de Student fuera debidamente apreciado. Como señaló McMullen, en el prólogo de la colección de artículos de Gosset, publicados en 1942, "Por un largo tiempo después de su descubrimiento y publicación, el uso de este test apenas salió de la cervecería Guinness".

El gran estadístico inglés Sir Ronald Aylmer Fisher ingresó a la Estación

Experimental de Rothamsted en 1919. Desde allí entregó una importante cantidad de conocimiento relacionado con el diseño de experimentos, contribuyendo a desarrollar técnicas que son consideradas claves para en la experimentación comparativa: El diseño experimental en bloques, que permite el control local del efecto introducido por factores no deseados, sobre las variables observadas. La aleatorización, que constituye una protección contra la introducción de factores impredecibles, en el experimento. El diseño factorial, para el estudio del efecto de varios factores, simultáneamente. Y el análisis de varianza, técnica de análisis de los resultados de la experimentación que permite separar las fuentes de variación, y así determinar el grado de influencia de cada factor. Estas técnicas, a excepción de la aleatorización, eran conocidas antes de Fisher, pero fue el quien logró una clara comprensión de ellas, e inició su uso en forma masiva. dio a conocer sus resultados en sucesivas publicaciones a partir de 1925, y hasta después de 1940, sobre métodos de experimentación agrícola.

Posteriormente estas técnicas fueron aplicadas a otras áreas no agrícolas, como la más reciente, la industrial. Esta se caracteriza por la utilización de experimentos factoriales fraccionados, que utiliza en forma óptima la información proveniente de un experimento desarrollado en forma parcial, debido al gran número de factores involucrados, y de la metodología del análisis de superficies de respuesta, un procedimiento fino para el estudio de los resultados experimentales.

Fisher desarrolló una teoría de estimación, aun en uso hoy día, basada en resumir los datos de un modo eficiente, que preserve la mayor cantidad de información contenida en ellos. Si se conoce la forma funcional de la ley de probabilidad que gobierna la población de donde provienen los datos, Fisher observó que la Función de Verosimilitud, la probabilidad de obtener la muestra dada, es un resumen de la información contenida en los datos. El método de maximizar la verosimilitud, provee entonces, el estimador más eficiente, que no puede ser mejorado, según su teoría.

Hacia fines de la década del 1950 existían pocos libros escritos sobre el tema del diseño de experimentos. Entre ellos, un aporte importante aparece en el de Oscar Kempthorne, y es el uso de matrices. Esta herramienta matemática, tan utilizada en la Estadística de hoy, sólo comenzó a utilizarse en Estadística a partir de esa década. Permitted un tratamiento más efectivo del análisis de varianza y del diseño de experimentos, en el contexto de un modelo lineal general. En esa misma década, con el acceso a los primeros computadores, que permitieron invertir matrices relativamente

grandes, lo que demanda un alto volumen de cálculo numérico, aparece la regresión múltiple, cuyo desarrollo todavía tiene lugar.

En el lema de regresión múltiple también incursionó George Snedecor, nacido en Estados Unidos en 1882, quien hizo grandes aportes al área del diseño de experimentos, en particular en aplicaciones a la agricultura. Hizo trabajo pionero en el uso de las tarjetas perforadas. Ligado al nombre de Snedecor, aparece el de William Cochran, nacido en 1909, en Escocia, quien hizo aportes al diseño de experimentos y a la teoría del muestreo. Ambos son coautores de un libro clásico sobre métodos estadísticos.

Con posterioridad al desarrollo del concepto de correlación, de Galton y de Pearson, problemas de clasificación en antropología y botánica dieron origen a coeficientes de similitud y a las funciones discriminantes del análisis multivariante. Dentro de esta misma rama de la estadística, pero en otra dirección, el estudio de las respuestas de los tests mentales dio origen a técnicas de reducción de dimensionalidad, es decir, el reemplazo de un gran número de variables correlacionadas, por un pequeño grupo de variables construidas a partir de las primeras, que contienen aproximadamente la misma cantidad de información. Entre ellas se encuentra el análisis factorial, que permite encontrar y cuantificar factores que influyen sobre las respuestas observadas. Las raíces del análisis factorial se encuentran en la psicología.

Otras técnicas del análisis multivariante se desarrollan como respuesta a problemas surgidos en otros campos, como el escalamiento multidimensional, el análisis de conglomerados, y el análisis de correspondencias, fuertemente relacionados con la mercadotecnia cuantitativa actual. Todas estas técnicas del análisis multivariante tienen un soporte matemático poderoso en el cálculo matricial, cuya utilización práctica en problemas de grandes volúmenes de observaciones y de variables se hace posible gracias a la aparición de los computadores. Entre quienes hicieron importantes aportes al análisis multivariante, se encuentra el estadounidense Harold Hotelling, estadístico, economista, con un entrenamiento inicial como periodista. Desarrolló la técnica de las componentes principales, que había sido iniciada por Karl Pearson. Hotelling hizo contribuciones importantes, además, al campo de la bioeconomía.

Otra rama de la estadística es la de los métodos no paramétricos, que traía con modelos estadísticos en que se hacen supuestos muy débiles sobre las distribuciones probabilísticas subyacentes. Un modelo paramétrico involucra el supuesto que esta

distribución pertenece a alguna familia, cuya forma general es conocida, pero de la cual se desconocen algunas características, o parámetros. Cuando no hay información que permita determinar una familia de distribuciones, es apropiado el uso de los métodos no paramétricos.

Entre quienes hicieron aportes de importancia a este campo, se encuentran Frank Wilcoxon, irlandés nacido en 1892, y Charles Spearman, inglés nacido en 1863. Wilcoxon recurrió a la simple idea de reemplazar los datos por sus rangos, al ordenarlos, sobre los cuales se pueden conocer propiedades distribucionales. Creó, de esta manera, una prueba basada en rangos, que hoy lleva su nombre. Esta idea inspiró el desarrollo de gran cantidad de otras pruebas, y del campo de la estadística no paramétrica, en general.

Charles Spearman sirvió en el ejército inglés, participando en la guerra de los Boers. Luego se retiró, para dedicarse a estudiar psicología, llegando a desempeñarse, ya tarde en su vida, como profesor de psicología en el University College, de Londres. Es conocido por sus contribuciones al análisis factorial, que se mencionó como una de las técnicas de la rama de la Estadística denominada análisis multivariante. También es conocido por sus investigaciones sobre la inteligencia. Estos intereses lo obligaron a estudiar estadística, llevándolo a desarrollar un coeficiente de correlación basado en rangos, que hoy lleva su nombre. El trabajo de Spearman ha sido desarrollado con posterioridad, desembocando en el análisis de varianza multivariante.

Al igual que Spearman, otros científicos sociales han entregado importantes contribuciones a la Estadística, lo que es indicación de su gran utilidad en el estudio de los fenómenos sociales. Entre ellos se cuentan L. Gutman y L.L. Thurstone, quienes se preocuparon de problemas de escalamiento, que consiste en transformar una característica medida en una escala conceptual, a una escala numérica. El escalamiento está fuertemente ligado al diseño y análisis de encuestas y tests.

Abraham Wald, húngaro que vivió entre 1902 y 1950, desarrolló la Teoría de Decisiones, entre 1939 y 1947, que constituye un modelo estadístico teórico, distinto a la escuela inferentista, de R. A. Fisher, dominante hasta entonces. Este último ve la estadística como un medio de hacer inferencias, de reducir la incertidumbre a través de la experimentación y la observación, o como un medio de resumir datos. La nueva escuela decisionista, de Wald, característica de los Estados Unidos, define la estadística como la ciencia de la toma de decisiones, bajo condiciones de incertidumbre.

Sí bien Wald desarrolló la teoría de decisiones en la forma actual, hubo antecesores que pensaron en términos de ella. Daniel Bernoulli, en 1730, introdujo nociones de utilidad, y de un espacio de acciones, elementos propios de la teoría de decisiones. También Laplace puede ser señalado como uno de los primeros decisionistas, al utilizar los cuatro elementos fundamentales de la teoría de decisiones: los estados de la naturaleza, la función de pérdida, las observaciones y el espacio de las acciones a tomar. Gauss también utilizó métodos que hoy se considerarían decisionistas. Lo hizo al contribuir al desarrollo del método de los mínimos cuadrados, que utiliza como criterio de decisión sobre un estimador, la minimización de una función de pérdida de tipo cuadrática. Gauss también compara el comportamiento de ésta con una función de pérdida función de valor absoluto. La teoría de pruebas de hipótesis desarrollada por Neyman y E. Pearson, entre 1928 y 1938, puede ser vista como un caso especial de la teoría de decisiones.

Otra contribución importante de Wald, es la de la inferencia secuencial, que toma decisiones que incluyen la opción de tomar más observaciones, cuando no hay evidencia categórica para tomar una decisión. Estos métodos son utilizados en la actualidad en el muestreo de aceptación, para el control de la calidad.

El estadístico George Box acuñó, en 1953, el término robustez, para designar los métodos estadísticos que procuran asegurar resultados aceptables, cuando no se cumplen los supuestos estándares en que se basan los métodos estadísticos regulares. Ya desde fines del siglo pasado hubo científicos que se preocuparon del tema. Se dieron cuenta de los peligros de hacer inferencias, cuando los datos aparecen contaminados con valores extraños, y llegaron a proponer modelos y estimadores robustos, como alternativas para estos casos. Sin embargo, no es sino a partir de la década de 1960, que este tópico es reconocido como un tema de investigación en Estadística. Y desde entonces, ha ido tomando importancia en forma progresiva. Entre los que le dieron el impulso a la estadística robusta, se encuentran Peter Huber y F. R. Hampel.

A partir de la Segunda Guerra Mundial, comienza la era de los computadores, que permitieron un acelerado desarrollo de la Estadística hacia regiones nuevas, caracterizadas por la aparición de técnicas cuya aplicación requiere de enormes cantidades de cálculos numéricos, imposibles de realizar con los medios existentes hasta entonces. Las dificultades de cálculo dejan de ser un impedimento, por lo que los modelos

estadísticos se vuelven más complejos. Los métodos de cálculo rápido, tan importantes en el pasado, quedan obsoletos. Paralelamente, aparece una gran cantidad de programas estadísticos envasados, fáciles de usar, que, tras ser alimentados con datos, producto de una investigación, entregan enormes volúmenes de resultados, que con frecuencia son erróneamente interpretados, y muchos de ellos son irrelevantes al propósito de la investigación. Pero bien utilizados, estos programas envasados permiten que las grandes masas de datos, productos de encuestas y censos, se vuelvan fáciles de administrar, y permiten que se mejore la calidad de ellos, al reducirse su manipulación.

Actualmente, la investigación en Estadística, cuyo resultado es la creación de nuevos métodos estadísticos, y una comprensión mejor de los métodos ya existentes, se apoya fuertemente en la computación. En el presente, el desarrollo de la Estadística parece ir junto con el desarrollo de la ciencia de la computación.

Bibliografía

T.W. Anderson: "An introduction to Multivariate Statistical Analysis", ed. John Wiley & sons, 1958.

Francisco Azorín - José Luis Sánchez-Crespo: "Métodos y Aplicaciones del Muestreo", ed. Alianza Editorial, 1986.

Peter Huber: "Robust Statistics", ed. John Wiley & sons, 1981.

Heinz Kohler: "Statistics for Business and Economics", ed. Scott, Foresman and Company, 1988.

K.V. Mardia - J.T.Kent - J.M. Bibby: "Multivariate Analysis", ed. Academic Press, 1978.

James Newman: "Sigma. El Mundo de las Matemáticas", ed. Grjaldo S.A., 1968.

D.B. Owen, editor: "On the history of Statistics and probability, ed. Marcel Decker, 1976.

Robert Steel - James Torrie: " Principles and Procedures of Statistics", ed. Mc Graw Hill, 1960.

Stephen Stigler: "The history of Statistics", ed. Harvad University Press, 1986.