

Contribuciones de Sir Ronald Fisher a la estadística genética

(E. A. Thompson, Departamento de Estadística, GN 22,
Universidad de Washington, Seattle, Washington 98195, U.S.A.)

Introducción

RESUMEN

Sir Ronald Fisher (1890-1962) fue profesor de genética y muchas de sus innovaciones estadísticas encontraron expresión en el desarrollo de metodologías en estadística genética. Sin embargo, mientras sus contribuciones en estadística matemática son fácilmente identificadas, en genética de poblaciones compartió su supremacía con Sewal Wright (1889-1988) y J. S. S. Haldane (1892-1965). Este documento muestra algunas de las mejores contribuciones de Fisher a las bases de la estadística genética, y sus interacciones con Wright y Haldane, los cuales contribuyeron al desarrollo del tema. Con la tecnología moderna, tanto la metodología estadística como la información genética están cambiando. No obstante, muchos de los trabajos de Fisher permanecen relevantes y pueden aún servir como una base para investigaciones futuras en el análisis estadístico de datos de DNA. El trabajo de este autor refleja su visión del papel de la estadística en la inferencia científica expresada en 1949.

ABSTRACT

Sir Ronald Fisher (1890-1962) was a genetics professor and many of his statistical innovations were shown in the development of genetic statistic methodology. However while his mathematical statistic contributions are easily identified, his supremacy in population genetics was shared with Sewal Wright (1889 - 1988) and J. S. S. Haldane (1892 - 1965). This document shows some of Fisher's best contributions to the genetic statistic bases and his interactions with Wright and Haldane, whom contributed to the development of the topic. With modern technology, statistical methodology as much as genetic information is changing. Nevertheless, many Fisher's works remain valid and can still serve as a base for future researches in DNA data statistical. This author's work reflects his vision about the statistic role in the scientific inference expressed in 1949.

No hay una amplia o urgente demanda por gente que defina métodos de prueba con el fin de mejorar la estadística matemática. Hay una urgente demanda por quienes entiendan la estadística genética, pero que sean capaces también de reconocer situaciones en el mundo real para las cuales la estadística es aplicable.

*Vicerrector de Educación UPB. Ex-cc. de Educación, Universidad Nacional de Colombia. Lic. en Matemáticas y Estadística, IPIC. Profesor visitante en UAGB. E-mail: jcuadros@telecom.net

En estadística, Sir Ronald A. Fisher es el primero en establecer las bases de buen arte de la teoría y práctica prevaletentes. Sus contribuciones son fáciles de identificar: desde el diseño de experimentos, ANOVA (análisis de varianza) y modelo lineal general, por medio de verosimilitud, máxima verosimilitud, información, consistencia y eficiencia, pasando por suficiencia, e intervalo de confianza, prueba de significancia, familia de exponenciales, familia de transformaciones y análisis discriminante lineal. Sin embargo, Fisher nunca fue profesor de estadística sino de genética. Una razón es que la mayor parte de sus aportes consisten precisamente en establecer la estructura de la inferencia estadística paramétrica relacionada con cuestiones genéticas, y modelos basados en análisis de verosimilitud como la principal estructura para inferencia genética; sus contribuciones a la estadística genética no pueden ser separadas de sus legados a la estadística general.

En genética de poblaciones, Fisher compartió su supremacía con J. B. S. Haldane y Sewal Wright, y es muy difícil señalar con precisión las ideas fundamentales, atribuibles solamente a Fisher. Estos tres fundadores de la genética poblaciones tuvieron carreras paralelas. Publicaron sus primeros documentos en 1912, siendo aún estudiantes: Haldane en Oxford, Fisher en Cambridge y Wright en Harvard. Haldane permaneció en la academia en Cambridge y en el University College (Londres) y viajó a India en 1957. Fisher trabajó como consultor estadístico en Rotamsted (1919 - 1933), retornando primero a la vida académica en el University College y posteriormente a Cambridge, como profesor de genética. Después de su retiro, viajó a Australia en 1959. Wright trabajó para la USDA de 1915 a 1925, viajó primero a la Universidad de Chicago. Se retiró a la Universidad de Wisconsin en 1955, donde a diferencia de los otros dos, vive 33 de sus más productivos años. Haldane y Fisher fueron por un tiempo colegas en el University College. En 1933 Fisher propuso un vínculo, al menos en la educación de estudiantes graduados (Box, 1978):

"Dado que los estudiantes preparados matemáticamente vienen a mí esperando obtener alguna clase de doctorado trabajando en mi departamento, sin saber nada, ninguna cosa, y no muy bien dispuestos a conocer alguna experimentación con material vivo, ¿puedo hacer que ellos se dediquen a lecturas en su departamento sobre teoría genética, así como así, un paso hacia la comprensión de las clases de razonamiento usada por los experimentadores?"

No obstante, Fisher, Haldane y Wright fueron competidores más que colaboradores, y no siempre fueron tolerantes con el trabajo de los otros. No hay duda que cada uno ganaba con las publicaciones de los demás, desarrollando ideas adicionales en respuesta a la crítica de los mismos. La esencia de sus contribuciones es muy diferente. Haldane tuvo diversos intereses; sus aportes son variados y hay menos de una perspectiva evolucionista para dar a su trabajo una síntesis global. Wright fue el más práctico, desarrollo teoría como respuesta a datos particulares. Fisher fue el teórico y tomó una visión más amplia de los datos, mientras la mayoría de su trabajo fue sugerido por hechos del mundo real, su interés estuvo en la construcción de una estructura sin la cual tales hechos no podrían ser tratados.

Algunas áreas de genética de poblaciones

El trabajo de genética que Fisher, Wright y Haldane encontraron como estudiantes, fueron los trabajos mendelianos y de variación cuantitativa, no enteramente aceptados. Hardy (1908) y Weinberg mostraron independientemente cómo la segregación mendeliana mantiene la variación de una forma tal que la teoría de la herencia no podría. Comenzó a comprenderse que los factores genéticos discretos se encontrarían bajo variación cuantitativa. El arreglo lineal de genes sobre cromosomas fue a ser aceptado por Sturtevant (1913) y las ideas de análisis de ligamento empezaron. Ideas acerca del muestreo de poblaciones (Bateson, 1909; Weinberg, 1912) fueron desarrolladas como estadísticas de particular importancia para el análisis genético.

En 1922, Fisher publicó tres importantes artículos: 1) presentó las bases para el futuro desarrollo de la genética de poblaciones teórica; 2) presentó las bases para el desarrollo de la Estadística Matemática; 3) demostró las ideas estadísticas teóricas del segundo, aplicándolas a la mayor área estadística del análisis genético -la construcción de mapas de ligamento genético-. Con esos tres documentos, Fisher establece su lugar como fundador de la genética de poblaciones, la estadística matemática y la estadística genética.

El trabajo de Fisher, Haldane y Wright fue adaptado a la biometría y a la genética, y presentó las bases del análisis cuantitativo de la evolución mendeliana. A pesar de su diferente perspectiva y sus teorías encontradas, sus contribuciones se mezclaron para dar forma a este análisis. Para ver cuán amplio es este consenso de teoría evolucionista, es instructivo revisar brevemente algunas de las áreas que ellos trataron; para evitar que este documento sea solamente una bibliografía, muchas referencias importantes deben ser omitidas. Felsenstein (1981) por ejemplo es una excelente fuente para la identificación de documentos no citados aquí.

Teoría evolucionista:

Mutación, selección y fijación de alelos raros

Una vez la selección natural llegó a aceptarse como la fuerza conductora del cambio evolucionista, el periodo propio para un extenso rango de estudios sobre la evolución de los sistemas genéticos se dio. En primer lugar, la teoría evolucionista se interesó por la evolución de los sistemas reproductivos y por la evolución de la dominancia; estos conceptos fundamentales fueron tratados repetidamente por Fisher, Haldane y Wright. Conjuntamente, presentaron una base perdurable para el más moderno trabajo en el área. Los nombres frecuentemente asociados con la primera teoría de la evolución de la recombinación son Muller (1932) y Haldane (1932), pero Fisher (1930) también hizo su contribución a esta discusión. En la misma época, él desarrolló la idea de un polimorfismo balanceado, un componente fundamental de la biología poblacional teórica. Sin embargo, Fisher es el mayor foco en esta época en la evolución de la dominancia, con su primer documento en 1922, y muchos otros entre 1928 y 1934. Si bien este tópico ha tenido menos impacto en la investigación reciente, él mismo pensó que su teoría de dominancia era de gran importancia y, por consiguiente, debía mencionarse. Además, este tópico ilustra mejor lo positivo de las interacciones adversas entre Fisher y Wright. El trabajo realizado por Fisher en sus artículos obtiene respuesta de Wright, retomando la discusión con un intercambio final en 1934. Haldane también participó en este debate. Este intercambio de críticas no sólo incitó desarrollos de la teoría, sino también exigió que cada adversario aclarara su pensamiento, un mayor beneficio para posteriores investigaciones.

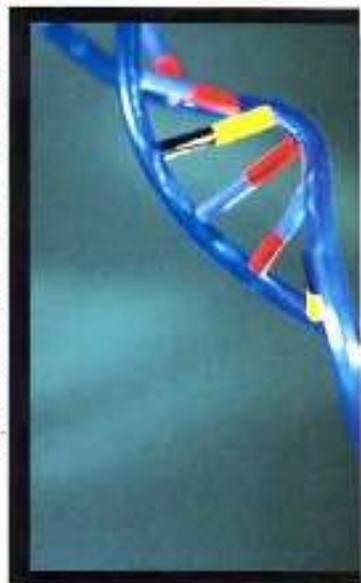
Mientras que la selección es la fuerza de la evolución, la mutación es la materia prima. La segunda área principal de la teoría evolucionista y la base de la mayoría de la biología poblacional teórica modela-

da y analizada, es también compartida por Fisher, Haldane y Wright. Esta es el área de la teoría comprometida con cambios de frecuencia de alelos bajo variación genética aleatoria y la última fijación o pérdida de alelos mutantes, resultando diferencias entre las especies. Fisher (1922) fue el primero en considerar tanto la ecuación para variación como la fijación de alelos mutantes; sin embargo, fue Haldane (1926) quien desarrolló la teoría de sustitución por alelos mutantes y más tarde la carga genética (1937); y fue Wright (1931) quien dio y desarrolló la primera formulación correcta de la ecuación de difusión de variación genética aleatoria. Kimura (1962-1964), el principal fomentador y adherente de la teoría neutral de la evolución. En 1958 empleo la edición de Fisher de 1930 como una de sus principales fuentes. Así, Fisher llega a tener un nombre frecuentemente asociado con la ecuación de difusión para variación genética. Sin embargo, en verdad, este es el único tópico en el cual Wright y Fisher estuvieron de acuerdo por un día; el "Modelo Wright-Fisher" (algunas veces también llamado "Modelo Fisher-Wright") es una piedra angular de la genética poblacional teórica.

Que Fisher fue el creador de "el teorema fundamental de la selección natural" es bien conocido, si el debate alrededor de este teorema continúa (Ewens, 1989). De mayor interés aquí son las diferentes perspectivas de selección -conducción propuestas por Fisher, por Haldane y por Wright-. Haldane, en virtud de su interés en mutación, investigó muchos casos de equilibrio-mutación- selección en una serie de documentos de 1926 a 1932, y desarrolló ideas de carga mutacional (Haldane, 1937). Wright, con sus teorías "paisajes adaptados" y "equilibrio cambiante" (1970), chocó con la visión de Fisher. La formulación de Wright -irónicamente para algunos, quienes opinan que su estructura poblacional e entrecruzamiento fueron de gran interés-, asumió el apareamiento aleatorio y la ausencia de interacción entre genes. La formulación de Fisher tomó cuenta del apareamiento e interacción, sobre lo cual comentaba mordazmente: "Yo nunca he escrito acerca de (media poblacional ajustada)... (la existencia de tal función potencia... no es una propiedad general de poblaciones naturales, sino que se origina únicamente en los casos restringidos y especiales de Wright... considera" (1958).

Este comentario es iluminación, por esto Fisher muestra su perspectiva global de modelos y matemáticas que debían ser aplicables al mundo real. Tanto Fisher (1918) como Wright (1917-1918) son bien conocidos (y para algunos quizás mejor conocidos) por sus contribuciones pioneras al análisis de variación cuantitativa y heredabilidad. Este no es el lugar para discutir su competencia por los derechos

a las bases del análisis de varianza. Los paralelos entre sus divisiones de la varianza total en sus documentos separados de 1918, y las diferentes trayectorias que indujeron a Wright a la teoría de coeficientes lineales (1912), y a Fisher a la tabla de análisis de varianza y la distribución F (1925), han sido discutidas en otra parte. Lo que es de interés aquí es que de nuevo Fisher no asumió el equilibrio de Hardy-Weinberg, e incluyó ordenadamente el apareamiento en sus modelos. El no discutió esos aspectos de su modelo detenidamente; más bien, consideró tales desviaciones de Hardy-Weinberg como otro caso cierto, asumiendo que cualquier modelo sería incluido.



Análisis genético humano;

determinación y análisis de ligamento

Una tercera razón por la cual es difícil limitar una discusión sobre Fisher a sus contribuciones estadísticas genéticas, es que muy poco de genética de poblaciones o de genética evolucionista es primariamente estadística, si bien tiene muchos elementos estadísticos dentro. Sin embargo, si uno fuera a preguntar a expertos en genética cuáles son las áreas de mayor demanda estadística, las respuestas deberían probablemente ser unánimes -determinación y análisis de ligamento-. Gran parte de lo que está hecho en ambas áreas fue trazado por el trabajo de Fisher de 1934 a 1936. En las áreas de análisis de ligamento y epidemiología genética humana han divergido de aquellas de evolución y teoría genética de poblaciones, pero los pioneros no ven razón para separar sus ideas. Fisher y Haldane trabajaron inicialmente en análisis de encadenamiento, y sobre determinación presentaron la base de muchas de las prácticas modernas en epidemiología genética. Sus contribuciones no son fácilmente separables. Si bien Fisher es mejor conocido por su trabajo sobre determinación, y Haldane y Smith (1947) por la extensión de las bases del análisis de ligamento, ambos fueron activos en dichas áreas durante los años 30.

La "determinación" ha venido a significar, en el



contexto de la estadística genética, el efecto del método de muestreo sobre la estimación de los parámetros de modelos genéticos. Mientras Fisher es frecuentemente citado por haber colocado las bases para corregir determinaciones, es interesante notar que éste no es el principal tema en el documento. Correcciones de determinaciones fueron consideradas anteriormente por Weinberg (1912), y la necesidad de tales correcciones fue considerada por Fisher como manifiesta:

Es común en estadística que la interpretación de un conjunto de datos requiere conocimiento de cómo éste fue obtenido. Igualmente es entendido que las conclusiones extraídas de resultados experimentales deben apoyarse en un conocimiento detallado del procedimiento experimental actualmente empleado. Sin embargo, en genética humana, los métodos estadísticos son algunas veces puestos descaradamente, y sus respectivas exigencias tomadas con entera negligencia de las condiciones de determinación.

Así, lo que más interesó a Fisher fue colocar el problema de muestrear en genética humana en un más amplio contexto estadístico el contexto de la inferencia máximo-verosímil, eficiencia e información de Fisher-. En efecto, Fisher se acreditó con Haldane, mostrando que los estimados de los parámetros considerados son estimadores máximo-verosímiles. Su

tema fundamental en este documento es sobre las eficiencias relativas de procedimiento de muestreo alternativo; esto permanece como un tópico de la investigación estadística genética corriente (Thompson, 1986).

El arreglo lineal de genes sobre cromosomas fue demostrado por Sturtevant (1913), por medio de un análisis estadístico de cálculos recombinados en *Drosophila*. Fisher (1922) mismo usó cálculos recombinados en cruzamientos diseñados para ilustrar sus desarrollos estadísticos teóricos. Hasta los documentos de Fisher de 1934 a 1936, revelan que él estaba poco interesado con la estimación máximo-verosímil que ha sido claramente desarrollado por Haldane (1934), cuyo documento incitó a Fisher al análisis. Mejor dicho, Fisher está de nuevo interesado en información y eficiencia y en la

pérdida de eficiencia debida a la estimación de otros parámetros genéticos (1934). El uso de máxima-verosimilitud y registro LOG (Log-cocientes de verosimilitud) es aún estándar en el análisis de ligamento humano; Fisher (1935) presentó la Infraestructura. Hay muchos intereses modernos en el uso del análisis de ligamento para planeamiento genético (Conneally 1989); Fisher (1934) previno sobre el más grande beneficio del planeamiento que debería resultar de estudios de ligamento. Activando investigaciones con nuevos y más eficientes métodos para estudiar la potencia de un estudio de encadenamiento (Ploughman, Boehnke, Lcinge y Matthyse, 1989).

Los tipos de sangre humana (RH);

un estudio con método científico

Los tipos de sangre humana proporcionaron (y para algunos tiende a proporcionar) una maravillosa y abundante oportunidad de información para aquellos interesados en variación mendeliana, y Fisher tuvo un perspicaz interés en desarrollos serológicos a comienzos de los 30's. Box (1978); él no estuvo únicamente interesado en teoría y formulaciones matemáticas. Fue en los 40's que Fisher (1947) hizo su mayor contribución al esclare-

cimiento de los tipos de sangre Rhesus, lo cual de nuevo muestra su visión de la inferencia científica (deducciones teóricas de datos del mundo real), la construcción de una estructura aplicable a las situaciones del mundo real e inferencia dentro del contexto de la estructura para una más amplia gama de problemas evolucionistas.

Fisher siguió considerando la disponibilidad de datos sobre frecuencias de segregación en familias un clásico análisis genético estadístico; sobre la base de la disponibilidad de datos, únicamente seis de ocho proporcionados por este sistema podrían ser validados, pero al instante después de eso un séptimo fue encontrado, y más tarde el octavo. Más importante aún, uno de los anticuerpos perdidos (anti-e) fue detectado; anti-d parecen no existir. Aquí realmente, está una "estructura aplicable al mundo real", lo cual hizo que se diera cuenta de la aparente caótica reacción Rhesus. Sin embargo, Fisher no se limitó a la construcción de la estructura; él consideró también las cuestiones evolucionistas reunidas por este sistema de tres loci muy estrechamente encadenados, y usó las frecuencias poblacionales de los ocho alelos presumiblemente originados de cada uno de los otros mediante muy extraordinarios eventos de recombinaciones para inferir el orden de los tres loci.

Así, en este ejemplo, nosotros tenemos a Fisher el matemático deductivo, Fisher el estadístico, y Fisher el teórico evolucionista. Este es la síntesis de esas tres perspectivas que caracterizan la singular contribución de Fisher. Él mismo sintió que este trabajo era un primer ejemplo de método científico (1947).

El advenimiento de los computadores

Una pregunta legítima es por qué Fisher y Haldane, habiendo reconocido la importancia del ligamento humano y del análisis de segregación, y presentando las bases de la teoría y la metodología, no continuaron con el tema más adelante. Hay quizás dos razones: una es que los datos para detallar mapas de encadenamiento genético humano no estuvieran disponibles. Únicamente con la llegada de señales (marcas) muy polimórficas de DNA ha sido posible mapear a gran escala encadenamiento, segregación y análisis de eslabonamiento. El advenimiento de los computadores más rápidos y más accesibles ha cambiado nuestra aproximación hacia la ciencia. Un trabajo de Fisher fue la derivación algebraica meticulosa y la extinción produci-

da por los cálculos manuales a tablas numéricas. Habría él dado la bienvenida a los computadores y la potencia enormemente incrementada para análisis de ligamento y el análisis de segregación que nosotros tenemos como resultado.

En efecto, él dio la bienvenida a los computadores. Haldane (1948) escribió un documento sobre Clinas -el patrón geográfico de frecuencias esperadas de alelos bajo selección y migración, cuando diferentes alelos son provechosos en diferentes regiones-. Fisher (1950), en un contexto similar tuvo una ecuación diferencial distinta para resolver, que podría no producir una solución analíticamente. La solución numérica proporcionada por el computador Cambridge EDSAC-T podría decirse, fue el primer uso publicado de un computador digital en biología matemática (Wiikes, 1975).

Cromosomas y datos de DNA

En los últimos años -puesto que Fisher muere en julio de 1962-, los datos han cambiado fuera de todo reconocimiento. Primero, los tipos de sangre fueron aumentando mediante polimorfismos de enzimas; luego, por polimorfismos de DNA, (Botstein et al., 1980); más tarde, por impresión digital de multilocus de DNA (Jeffreys, Willson y Thein, 1985), y ahora por polimorfismos de un único locus VNTR (Nakamura et al., 1987). La información secuencial de DNA está disponible actualmente en (mapa genético). En principio, los nuevos polimorfismos no son muy diferentes; del mismo modo los tipos de sangre proporcionan señales mendelianas cuantitativas que pueden ser usadas para asignar diferenciación poblacional y variación individual. La diferencia se encuentra en el hecho que la densidad de la información sobre los genomas está ordenada en magnitud creciente. En lugar de esforzarse para encadenar unos pocos pares de sistemas de tipos de sangre, y quizás asignárselos a uno de los 22 pares de autosomas, los expertos en genética pueden ahora ensayar a ordenar 50 señales de DNA, todas condicionadas por estar situadas en una pequeña región cromosomal. Análisis basados en modelos oligogénicos no son aplicables a tales datos; una nueva estructura en la cual el cromosoma es tratado como una continuación de información genética es necesario. En efecto, semejante estructura fue proporcionada por Fisher (1949) en el contexto de sus estudios de intracruzamiento.

En su serie de cinco documentos pioneros en 1921, Wright (1921) coloca la base original de esta área; las

futuras generaciones de contribuciones provinieron de Cotterman (1940) y Malecot (1948), más que de Fisher o de Haldane. El trabajo de Wright sobre estructura poblacional condujo también a su teoría de aislamiento por distancia (1943), con la cual a pesar del trabajo de Fisher sobre la ecuación de difusión (1922) y sobre clínicas (1950), hizo Wright la mayor contribución a esta área. Sin embargo, el trabajo de Wright en entrecruzamiento y estructura poblacional e entrecruzamiento fue en el contexto de un único (o algunas veces múltiple, pero segregado independientemente) loci. Haldane tuvo un interés en entrecruzamiento en el contexto de letales recesivos, y también, a través de esto, en entrecruzamiento de loci encadenado (Cf. Haldane, 1949). La principal contribución de Fisher en esta área es un pequeño manual sobre la teoría de entrecruzamiento (1949). La mayor parte del texto consiste en derivaciones algebraicas de la proporción de pérdida de variación genética en un único locus, bajo sistemas alternativos de apareamiento, en meticuloso detalle. Sin embargo, hay una sección (poco conocida) del texto que puede llegar a ser apreciada como una de las más grandes contribuciones de Fisher a la estadística genética, pues proporciona la estructura necesaria para el análisis de una cadena de cromosomas continua. Antes que considerar la descendencia e entrecruzamiento de genes en puntos (loci) sobre un cromosoma, Fisher (1949) desarrolló la idea de puntos de empalme sobre los cromosomas, en los cuales las recombinaciones han ocurrido proporcionando un punto de reunión entre segmentos de cromosoma de diferente origen ancestral. Considerando el destino de esos empalmes, los procesos de su creación, pérdida o fijación, es posible considerar entrecruzamiento y pérdida de variación genética no sólo en un simple loci, sino sobre la longitud del cromosoma. Fisher desarrolló su estructura teórica, reconociendo que los cromosomas eran hechos del mundo real, y que la genética de poblaciones requería una estructura para describir la herencia y descendencia de segmentos de cromosoma. Interesado, él también simuló el proceso (a mano), indicativo tanto de un modelo bien definido como de su interés en los resultados de corto plazo que él no podría derivar analíticamente. Esta teoría y resultados adelantarian su tiempo, y con unas pocas excepciones (Cf. Robertson, 1970; Frankini, 1977; Donnelly, 1983) han sido ignorados.

Fisher parece haber sido asignado a su historia. En septiembre de 1961, en su visita final a Stanford en su camino de regreso a Australia, él sugirió la teoría de empalmes como el tópico para su seminario (Cf. W. F. Bodmer; Fisher Memorial Lecture, Cambridge 1990). Él fue persuadido a hablar sobre otro tópico; los empalmes no fueron entonces considerados como un tópico prometedor para atraer a la audiencia. Fisher

desarrolló su teoría de empalmes como una teoría dirigida a una "situación en el mundo real" -el hecho de la continuidad del material genético-; él seguramente habría dado la bienvenida a los modernos datos genéticos que hacen esta teoría aplicable estadísticamente.

Con el transcurso de los años, el aspecto del trabajo de Fisher que ha tenido mayor impacto en la investigación corriente en estadística genética ha cambiado continuamente. Durante los años 70's, la teoría evolucionista básica fue silenciosamente desarrollada. Fisher (1930) describió la teoría y, en conjunto, la escena para desarrollos futuros. Durante los años 30's, las ideas de análisis de encadenamiento fueron iniciadas y ha permanecido en uso desde entonces. Durante los 40's, los lipos de sangre llegan a ser una área importante de investigación. Los 50's vieron renovada la actividad en el área de genética cuantitativa, particularmente en las áreas de crianza de plantas y animales; en los 60's, con la enorme expansión de la biología poblacional matemática. El trabajo de Fisher sobre teoría evolucionista recibió renovada atención. En los 70's, con el crecimiento del trabajo en análisis epidemiológico genético humano de complejas características, Fisher contribuyó a que la teoría de determinación llegara a ser bien conocida. En los 80's tuvo lugar la década del análisis. Las bases estadísticas fueron presentadas por Fisher y Haldane entre 1934 y 1936. En los 90's, polimorfismos del DNA y secuencias de DNA llegaron a ser los datos de selección para dirigir cuestiones de epidemiología genética y biología evolucionista: a través de su historia de empalmes, la más grande contribución de Fisher a estadística genética que aún puede ser usada.

BIBLIOGRAFÍA

- +AFIFI, A.A., y CLARK, V. (1984). Computer-aided multivariate analysis. Lifetime Learning Publications, Belmont California.
- +AGRESTI, A. (1984). Analysis of ordinal categorical data. John Wiley & Sons, New York.
- +AGRESTI, A. (1990). Categorical data analysis. John Wiley & Sons, New York.
- +ANILGREN, A. (1990). Statistics in Project 2061 Curriculum Development, Proceedings of the Third International Conference on Teaching Statistics 1, 36-38, Dunedin, Nueva Zelanda.
- +ALDENBERGER, M.S., y BLASHFIELD, R.K. (1984). Cluster analysis. Sage Publications, Beverly Hills, California.
- +ALDRICH, J.H., y NELSON, F.D. (1984). Linear probability, logit and probit models. Sage Publications, Beverly Hills, California.
- +ANDERSON, E.B. (1989). The statistical analysis of categorical data. Springer Verlag, New York.
- +ARNOLD, S.F. (1981). The theory of linear models and multivariate analysis. John Wiley & Sons, New York.
- +BENZECRI, J.P. (1992). Correspondence analysis handbook. Marcel Dekker, Inc., New York.
- +BERENSON, M.L., y LEVINE, D.M. (2000). Basic Business Statistics, concepts and applications. Prentice-Hall, Englewood Cliffs, New Jersey.
- +BREWER, J. K. (1972). On the power of statistical tests in the American Education Research Journal. American Education Research Journal.