

## Nuevas tendencias de la Estadística en el siglo XXI\*

José A. Cristóbal

Departamento de Métodos Estadísticos

Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza

La llegada del nuevo milenio ha impulsado un nuevo espíritu en la investigación tanto para científicos como para humanistas. En particular, todo apunta hacia el desarrollo de una nueva cultura intelectual donde el trabajo más profundo de los científicos se hace más accesible al público en general. En palabras de John Brockman, filósofo norteamericano y autor de numerosos libros sobre ciencia de vanguardia: “La aparición de esta nueva cultura es una evidencia de un gran anhelo universal, un deseo para la nueva e importante idea que dirige nuestros tiempos: un desarrollo revolucionario en biología molecular, ingeniería genética, nanotecnología, inteligencia artificial, teoría del caos, redes neuronales, biodiversidad, sistemas expertos, autómatas celulares y realidad virtual, entre otros”. Al menos, términos tales como ADN, wavelets, fractal, ciberespacio, etc. son ahora frecuentes en nuestro vocabulario.

En lo que se refiere a la Estadística, el británico Peter Green ha resumido muy bien esta omnipresencia: “Como todos nosotros reconocemos, la Estadística es una disciplina extraordinaria. Se ha extendido dentro de la administración y la industria, en la vida pública y en muchas otras áreas. A su vez, esas interacciones han moldeado profundamente esta materia. Su concepción adopta una amplia variedad de aspectos: fundamentos filosóficos, teoría matemática, principios de inferencia, diseño, recogida de datos, técnicas, computación, construcción de modelos, etc. Y crucialmente, el comportamiento de sus interacciones con el resto del mundo es parte de su propio objetivo”.

Pero “algunos usuarios de métodos estadísticos, ven aparentemente a la Estadística como algo estático, un armario lleno de técnicas y buenas prácticas para ser consultados sólo ocasionalmente, y sin mucho entusiasmo ni compromiso”. De hecho, en los foros y reuniones científicas en donde la Estadística no es el objetivo prioritario del acontecimiento,

---

\*Conferencia leída en el Aula Magna de la Facultad de Ciencias de la Universidad de Zaragoza, el 15 de noviembre de 2006, con motivo de la festividad de San Alberto Magno.

se puede observar constantemente el uso de términos tales como proceso generador, análisis de datos, predicción, estimación, hipótesis, significación, etc., pero la Estadística como disciplina suele ser raramente reconocida.

En un mundo en el que “*información*” se ha convertido a la vez en una moneda de cambio y un producto global, en el cual “*incertidumbre*” es un término en creciente utilización y su impacto más ampliamente apreciado, y en el que gobierna la “*cuantificación*”, es asombroso que una disciplina cuyo núcleo central es la *cuantificación de la información bajo incertidumbre*, no sea debidamente valorada.

*¡Claramente, tenemos un problema de imagen!*

Me atrevo a apuntar la necesidad de un cambio de actitud por parte de los estadísticos para conocer las necesidades que se despliegan en la ciencia del siglo XXI.

Es mi propósito, en estas breves palabras, comentar algunos aspectos de nuestro pasado estadístico, examinar la situación en que nos encontramos en el presente y especular con nuestro futuro.

En primer lugar, el desarrollo de la Estadística ha estado condicionado por el extenso entorno en el que opera: por aplicaciones en las ciencias, las ciencias sociales, la medicina, la ingeniería y el mundo empresarial, por la aparición de nuevos tipos de datos que están demandando interpretación y análisis, y por el rápido avance de la tecnología de la computación. Esta reflexión no es nueva, pero lo que parece no tener precedente es el rango de aplicaciones y nuevos tipos de datos que están empujando el desarrollo de la Estadística. En los siglos XIX y XX, el análisis estadístico estuvo en gran parte dirigido a las aplicaciones en un pequeño número de áreas (astronomía, estadística oficial, agricultura). Pero en la segunda mitad del siglo XX, la Estadística se ha convertido en una parte central de muchas disciplinas que analizan datos numéricos, (y a veces, incluso no numéricos), y una buena parte de la investigación estadística ha sido consecuencia de la demanda de nuevos métodos desde estas disciplinas externas.

Por otra parte, es un hecho claro que la revolución de los computadores ha transformado la Estadística, la cual se ha construido en gran parte sobre fundamentos matemáticos, pero en los últimos 30 años la computación de alta velocidad se ha convertido en una piedra angular. Ha hecho posible nuevos tipos de análisis y modelos que previamente no sólo eran impracticables, sino incluso impensables. Estas técnicas abarcan desde el primer software interactivo, tal como GLIM en los años 70, pasando por el bootstrap y el lenguaje S, que permitió fácil exploración visual de los datos en los años 80, hasta la revolución bayesiana de los 90, hecha posible gracias a los métodos MCMC, acrónimo de

Markov Chain Monte Carlo (por cierto, este año celebramos el centésimo quincuagésimo aniversario del nacimiento de Markov).

En general, quizá podemos decir que la investigación científica está dedicada a la búsqueda de teorías o hipótesis que explican o predicen un fenómeno. ¿Cómo puede ayudar la Estadística a conocer si una hipótesis es cierta o no? Es bien conocido que hay varias escuelas dentro de la Estadística que propugnan enfoques diferentes para resolver un mismo problema. El gran estadístico hindú, Calyampudi Rao plasmó magníficamente esta situación, al describir una conversación imaginaria entre Albert Einstein y algunos estadísticos, líderes de esas filosofías:

*Einstein:* Tengo una nueva teoría para explicar algunos fenómenos naturales. ¿Puede ayudarme la Estadística a contrastarla?

*Neyman y Pearson:* Albert, tienes que hacer tu experimento, darnos tus datos y decirnos qué posibles alternativas hay a tu teoría. Nosotros podemos revelarte el método más potente de verificarla.

*Einstein:* ¡Teorías alternativas! Puede haberlas, pero no conozco ninguna.

*Fisher:* Puedo darte el diseño de un experimento perfecto para llevarla a cabo. Los resultados pueden rechazar tu teoría si es errónea, aunque no pueden confirmarla si es cierta.

*Wald y Wolfowitz:* Podemos ayudarte a decidir eligiendo la acción que más te conviene, pero necesitamos que nos des una valoración cuantitativa del coste que ocasionará tanto la aceptación como el rechazo de tu teoría.

Por supuesto, sería conveniente añadir a este diálogo la posibilidad de un enfoque bayesiano, puesto en boca de alguno de los grandes pioneros de este tratamiento (como por ejemplo Bruno De Finetti, del cual también celebramos este año el centésimo aniversario de su nacimiento).

Dada la gran variedad de opciones, el estadístico inglés David Cox ofrece su propia perspectiva sobre estas cuestiones: “Mi punto de vista personal es ecléctico, consistente en tomar lo que es valioso de los diferentes enfoques. Por supuesto esto puede degenerar en una cobardía vagamente intelectual, pero la defensa de mi decisión radica en tener en cuenta el test fundamental de ver lo que hay de relevante y fructífero en las aplicaciones. La extraordinariamente rica variedad de aplicaciones de la Estadística apunta en contra de tomar un enfoque simple”.

Ante tal panorama, cabe preguntarse: ¿hacia dónde se dirige el área de la Estadística en el nuevo milenio? Al estadístico irlandés Adrian Raftery le parece seguro que se realizarán progresos teóricos para analizar nuevos tipos de datos, con el consiguiente

desarrollo de técnicas computacionales para hacerlo posible. Un ejemplo lo constituyen los datos de expresión genética, área en el que los estadísticos se han introducido rápidamente. La minería de datos constituye otro ejemplo; su estudio comenzó como un análisis de datos codificados, aunque los estadísticos se han incorporado en este campo de modo más lento. Por otra parte, un área donde la Estadística ha estado ausente en el pasado, pero donde se prevé que las nuevas teorías y potencia computacional permitan una relevante contribución, es el análisis de simulación de modelos mecánicos, donde domina el esfuerzo científico de muchas disciplinas para excluir los modelos estadísticos más convencionales.

Aunque el abanico de posibilidades sobre las futuras direcciones de la Estadística es enorme, voy a fijarme en tres grandes secciones sobre las que hay un cierto consenso en considerar como más importantes: *la Estadística en las Ciencias de la Vida, la Estadística en la Empresa y las Ciencias Sociales, y la Estadística en las Ciencias Físicas y la Ingeniería.*

Por supuesto, hay una gran diversidad y amplitud de técnicas en esos tres contextos, pero diversos temas son a veces recurrentes, subrayando la unidad del campo de la Estadística. Como ejemplo, consideremos el análisis de procesos puntuales definidos por las veces que ocurre a lo largo del tiempo un determinado fenómeno, tal como una muerte, una salida de un empleo o el fallo de una máquina. En las ciencias de la salud, esto se llama análisis de supervivencia; en las ciencias sociales, análisis temporal (o histórico) de sucesos, y en la ingeniería se denomina teoría de la fiabilidad. La estrategia de análisis subyacente es la misma en las tres áreas: el análisis de la función de fallo y el desarrollo de modelos para este concepto, como el célebre modelo de riesgo proporcional de Cox.

#### *I.- Estadística en las Ciencias de la vida*

Paso a considerar el primero de estos contextos, que es el de la Estadística en las Ciencias de la vida, donde el desarrollo y aplicaciones de un conjunto coherente y comprensivo de métodos para analizar datos es quizá el mayor logro de la Estadística en la segunda mitad del siglo XX. Esto ha llevado al desarrollo de la Bioestadística, la cual es una próspera disciplina en sí misma, mientras permanece como una parte integral de la amplia profesión estadística.

Varias materias comunes aparecen en este campo, pero destacaré tres importantes metodologías: el análisis causal, el análisis de supervivencia y los modelos jerárquicos, así como un rico conjunto de aplicaciones.

El análisis causal utilizando el concepto de efecto de un tratamiento fue introducido por Neyman para experimentos en el área de la agricultura, y el método de aleatorización

de Fisher le dio la base física para realizar inferencias causales. El epidemiólogo inglés Sir Austin Hill (pionero en el descubrimiento de la relación del consumo de tabaco con el cáncer de pulmón), introdujo esos principios en el campo de la medicina dando lugar a los ensayos clínicos aleatorizados, doble-ciego y controlados con placebo. Como David Harrington (célebre investigador en distintos aspectos del cáncer y profesor de Bioestadística en Harvard) afirmó, éste fue posiblemente el más importante avance científico en medicina durante el siglo XX.

El impacto del análisis de supervivencia ha sido inmenso: se han discutido métodos de supervivencia para monitorización de la fertilidad de poblaciones. Se han descrito modelos para el estudio del cáncer sobre la base del análisis de tiempos de vida en ensayos con animales. El estudio de los efectos del SIDA ha estimulado el desarrollo de metodologías para tratar con estructuras de datos cada vez más complicados que pueden estar sujetos a truncamiento y censura. Por otra parte, los métodos de supervivencia multivariante se han revelado de una gran utilidad y aplicación en epidemiología genética.

Pero este amplio conjunto de aplicaciones todavía representan sólo una pequeña muestra del total. El estimador límite producto de Kaplan y Meier, el test del logaritmo del rango de Mantel, y el modelo de regresión de riesgo proporcional son las técnicas indispensables de un gran plantel de estadísticos trabajando en ensayos clínicos en la industria, el gobierno y en el mundo académico. El hecho de que David Cox recibiera el premio *General Motors* de 1990 para investigación sobre cáncer clínico subraya el impacto enormemente beneficioso de su trabajo en la medicina clínica.

La medicina preventiva se ha visto no menos afectada por los conceptos y métodos del análisis de supervivencia. La medida epidemiológica clave, la razón de incidencia, está firmemente arraigada en la tradición de varios siglos de las tablas de vida, mientras que el más reciente concepto de riesgo relativo está basado en un ratio de tales índices. El modelo de riesgo proporcional dota de los fundamentos matemáticos a los métodos epidemiológicos clásicos de estimación de riesgo relativo.

En cuanto a los modelos jerárquicos, aunque se pueden implementar utilizando sólo el modelo mixto de ecuaciones y la estimación de componentes de la varianza por máxima verosimilitud restringida, las ventajas obtenidas al realizar un enfoque totalmente bayesiano son aparentemente mayores. Los cálculos MCMC, además, son esenciales para ajustar modelos con un gran número de efectos aleatorios.

Por otra parte, el método Bayes de selección de modelos se ha mostrado como muy útil en epidemiología cuando hay factores de exposición múltiple, como sucede, por ejemplo, al examinar la relación entre polución atmosférica y mortalidad. En la actualidad,

se están desarrollando con rapidez aplicaciones de MCMC para predicción espacial en problemas de datos de monitorización medioambiental (como por ejemplo la estimación de niveles de ciertos contaminantes en lugares que no pueden ser medidos). El motivo es que estas técnicas no tienen el inconveniente de los métodos geoestadísticos tradicionales, que necesitan realizar fuertes hipótesis sobre la estructura espacial.

No obstante, los bayesianos no están solos en su uso de MCMC y otros procedimientos computacionalmente intensivos. El bootstrap de Efron ha tenido también un impacto espectacular tanto en la teoría como en la práctica de la Estadística. Los modelos de captura-recaptura se han aplicado, entre otros contextos, a la estimación de parámetros demográficos de poblaciones de animales en libertad, tales como tamaño de población, supervivencia, migración, etc. De hecho, la continua interacción entre expertos en ecosistemas y estadísticos ha sido crucial para el desarrollo de dichos modelos de captura-recaptura (atrás quedaron ya los años en que una población animal se estudiaba a partir de una foto fija en el tiempo. Ahora sabemos que los datos así recogidos están sesgados y también sabemos cómo modificar la metodología estadística estándar para analizarlos).

Pero quizá el área más activa de la ciencia en este momento lo constituye el estudio del genoma. Cuatro líneas de investigación son las más relevantes en este campo: las dos más establecidas de localización de genes y análisis de secuencias y las dos más actual y rápidamente extendidas (en donde la Estadística interviene de un modo más importante), que son la predicción de estructura de proteínas y el análisis de datos de expresión de genes. Y es que, mientras en el siglo XX los esfuerzos estuvieron dedicados al análisis individual de los genes, en este nuevo siglo se plantea como objetivo el conocimiento conjunto de la naturaleza completa del genoma, a través de los datos genéticos.

Este pequeño listado que acabamos de hacer representa tan sólo una muestra de los fascinantes problemas estadísticos que aparecen en las ciencias médicas y de la vida, de las cruciales contribuciones hechas por estadísticos en este área, y de los conceptos y técnicas estadísticas que han conducido con éxito a su resolución. Pero sirven para confirmar que la Estadística del siglo XXI estará fuertemente influenciada por los desarrollos revolucionarios de la tecnología, particularmente en las ciencias biomédicas, y por la posibilidad de acceder a nuevos enormes volúmenes de datos geográficos, medioambientales, moleculares, etc.

Podemos acabar este punto mencionando el refuerzo de actividades de colaboración dado en el informe de 2002 de la National Science Foundation, que realiza el siguiente alegato: “La gran cantidad de datos producida por experimentos biológicos modernos y la variabilidad en la respuesta humana a la intervención médica produce una creciente demanda de estadísticos que pueden colaborar con biólogos e inventar nuevos métodos para

conducir diseños experimentales y análisis de datos biológicos. Existe así una necesidad de coordinar fondos a largo plazo para proyectos interdisciplinarios tales que el estadístico pueda ser capaz de desarrollar el entendimiento científico tan vital a la verdadera colaboración”.

## *II.- Estadística en la Empresa y las Ciencias Sociales*

En cuanto al segundo contexto de análisis, el de la Estadística en la Empresa y las Ciencias Sociales, hay que destacar la importancia histórica y el papel del pensamiento y la metodología estadística en la evolución de los estudios en este área, así como la razonable seguridad de la consecución de continuas innovaciones en las intersecciones de la Estadística con esos campos de conocimiento. Los aspectos metodológicos más habituales de la Estadística en este área lo constituyen el análisis de series de tiempo y la predicción, las tablas de contingencia, y la inferencia causal. Por otra parte, atendiendo a las disciplinas en donde se aplica, la gama es extensa y variada, pero podemos citar como más representativas, las áreas de finanzas y el marketing, las ciencias políticas, la psicología, la sociología, la demografía y el derecho.

Por supuesto, no se pretende realizar aquí una revisión enciclopédica de todos estos temas y, ciertamente, se podrían haber incluido muchas otras técnicas adicionales de haber suficiente espacio y tiempo. Dicho esto, parece lógico destacar algunos de los tópicos que son recurrentes en todas las ramas científicas anteriormente citadas, como son el análisis de datos de panel, los modelos de ecuaciones estructurales, el análisis factorial y otras técnicas multivariantes, el análisis espacial, etc.

Puede parecer extraño que no haya comenzado este punto destacando el área de la Economía, pero su interrelación con la Estadística es tan fuerte que para explicar el papel de la Estadística en la Economía sería esencialmente necesario hacer una revisión de prácticamente todo el campo de la propia Estadística. En lugar de ello, me dedicaré a soslayar algunas de las numerosas referencias a los papeles que la Economía y los economistas han jugado en el desarrollo de metodología estadística dentro del campo de la administración de empresas y las ciencias sociales.

La Estadística y los mercados financieros han estado desde el comienzo fuertemente relacionados, pero en las últimas décadas, muchas de esas conexiones han llegado a ser una parte del núcleo de las finanzas modernas, y en la actualidad es imposible entender el trabajo sobre mercados financieros sin su ayuda. A este respecto, cabe destacar que el Premio Nobel de Economía de 2003 recayó en Robert Engle (compartido con Clive Granger), por “el empleo de métodos estadísticos en series económicas de tiempo, y

en concreto por los métodos de análisis de series temporales económicas con volatilidad temporal variable”.

Hay muchos ejemplos en los que la Estadística se ha convertido en indispensable para realizar análisis financieros. Así, el análisis multivariante, la entropía y otros conceptos de teoría de la información, se utilizan para resolver, con ciertas propiedades de optimalidad asintótica, el problema de la selección de la cartera de valores. Los métodos no paramétricos, como la suavización local y el algoritmo de remuestreo bootstrap son ya lugar común en la estimación y evaluación de modelos financieros, muchos de los cuales son altamente no lineales y basados en ingente cantidad de datos. Las wavelets, las máquinas de soporte vectorial y otros modelos de series de tiempo no lineales se han aplicado con éxito a los problemas de predicción financiera y administración del riesgo.

De igual modo, no es sorprendente que gran parte del desarrollo metodológico sobre los estudios empíricos de marketing tenga un cierto sabor estadístico, o que la literatura econométrica figure de modo prominente en cualquier revisión sobre series de tiempo y predicción. En cualquier caso, en los últimos años se ha originado una explosión en la cantidad y variedad de datos que son factibles para los investigadores de mercados. En particular los datos que provienen de comercio electrónico y marketing directo han crecido de modo exponencial y la tecnología basada en la web ha disminuido espectacularmente el coste de la obtención de datos de demanda. Ello ha impulsado la aplicación de modelos jerárquicos bayesianos, entre otros, que constituyen un planteamiento natural para realizar inferencias en este contexto.

En lo que respecta a la utilización de técnicas estadísticas en el área jurídica, se pueden destacar tres grandes ramas de aplicación: a) utilización de análisis empírico científico para hacer prevalecer un caso en un litigio (en este sentido, el ejemplo más importante se refiere a la identificación forense —recordemos la masiva utilización en la actualidad del ADN para estimar modelos probabilísticos de identificación de sujetos—), b) utilización de análisis empírico social en casos individuales (como puede ser el análisis de casos de discriminación —por ejemplo en el empleo—), y c) utilización de métodos empíricos para describir cómo opera el sistema legal (cuya consecuencia más importante puede ser la creación de ciertas políticas públicas —en este campo la bases de datos suelen ser fuertemente sesgadas y afectadas por la no-respuesta o la no-observación, por lo que se necesitan técnicas estadísticas no estándar—).

Por otra parte, en lo que se refiere a la utilización de tratamientos estadísticos en las ciencias políticas cuantitativas, puede citarse el final de la Segunda guerra Mundial como la época de la revolución de las ciencias del comportamiento en la metodología política, aunque ya se habían realizado serios intentos desde finales del siglo XIX. En realidad los



científicos políticos encuentran el enfoque econométrico más atractivo que el estadístico, porque, simplificando, éste último tiene como fin primario la evaluación de teorías (y por lo tanto, están más interesados en saber si un parámetro de un modelo es grande o pequeño, en lugar de la relación exacta que hay entre las variables). Es conocido el aforismo: si el modelo y los datos no son congruentes, el estadístico cuestiona el modelo, mientras que el economista cuestiona los datos. Es mi experiencia que ambas partes están de acuerdo con esto, y nadie se siente ofendido en esta descripción.

Por lo tanto, aunque los tipos de cuestiones que aparecen en los variados campos que se originan en la interconexión de la Estadística con las ciencias sociales varían en alguna medida, es claro que el pensamiento y la metodología estadística es una parte central del avance de nuestro entendimiento del comportamiento y las interacciones humanas. Las oportunidades presentadas por las nuevas y evolucionadas tecnologías para recoger más y mejores datos son abundantes, y sin duda continuarán motivando nuevas investigaciones y aplicaciones estadísticas durante muchos años. Un atractivo de la Estadística ha sido siempre su amplia aplicabilidad a interesantes e importantes problemas, y la anterior revisión es una muestra de las numerosas e intelectualmente desafiantes oportunidades para hacer valiosas contribuciones en distintas áreas.

### *III.- Estadística en las Ciencias Físicas y la Ingeniería*

El último de los contextos que analizamos es el de la Estadística en las Ciencias Físicas y la Ingeniería, campo en el cual tienen su raíz muchos de los problemas reales a los que se enfrentaron los estadísticos a partir del siglo XIX. Así, por ejemplo, Box estudió diseños de superficies de respuesta trabajando con ingenieros químicos, Tukey desarrolló el análisis exploratorio de datos trabajando con ingenieros de telecomunicación, y Wald llevó a cabo la teoría de los tests de hipótesis secuenciales trabajando con ingenieros militares. Estos estadísticos tenían a la vez un fuerte sentido de lo que era importante en la correspondiente área de aplicación, así como un gran conocimiento de la metodología estadística que se necesitaba.

Las aportaciones de la Estadística en este área se desarrollan en muy variados campos de aplicación, comenzando con ciencias de la tierra y continuando con telecomunicación, control de calidad, monitorización y producción avanzada. Bajo una mirada superficial, esas materias tienen poco en común, pero comparten algunas similitudes profundas. Muchas de esas áreas, por ejemplo, se enfrentan a nuevas oportunidades y desafíos a causa de nuestra creciente habilidad para recoger extensas cantidades de datos. Cada vez con mayor frecuencia, la unidad de tamaño muestral en las ciencias físicas e ingeniería no es el número de observaciones, sino el número de gigabytes de espacio necesitado para almacenar los datos. Pero a pesar de los tremendos avances en potencia computacional,

el procesamiento de bases de datos tan extensas puede entrañar una gran dificultad, y la visualización, exploración y construcción de modelos para los datos puede ser aún más complicado. No obstante, una ventaja de trabajar con físicos e ingenieros es que muchos de ellos tienen años de experiencia diseñando sistemas de recogida, proceso y modelización de conjuntos masivos de datos, y los estadísticos pueden aprender de su experiencia. Lo importante es que muchos de estos avances están animando a los estadísticos a colaborar con especialistas de cualquier área, incluso de aquellos campos en las que esta colaboración era totalmente impensable hace sólo unos pocos años.

Una reflexión a tener en cuenta es que los modelos estadísticos por sí solos pueden ser insuficientes para resolver algunos problemas. Es necesario que incorporen los conocimientos de una teoría científica a la aleatorización. Por ejemplo, el problema de la predicción de terremotos, que actualmente está minando la credibilidad de los científicos que trabajan en este campo, requiere tanto una adecuada teoría de génesis y evolución sísmica, como la construcción de modelos estadísticos que respetan las leyes geofísicas y sean apropiados para datos altamente agrupados. Recordemos a Box: “todos los modelos son falsos, aunque algunos son útiles”.

Otro tanto puede decirse cuando se analiza el flujo de datos a través de Internet. Cuando se supervisa la bibliografía sobre este tema, un adjetivo emerge de manera recurrente para describir este tráfico: *feroz*. Las bases de datos son enormes, la topología de la red es vasta, las propiedades estadísticas son complejas, y el mecanismo de ingeniería es intrincado e introduce feedback en el tráfico. Para que los modelos estadísticos alcancen aquí el éxito necesario, es preciso que tengan en cuenta el conocimiento del entorno que rodea este complejo mecanismo.

Incluso en áreas de ingeniería en las que la incorporación de la Estadística tiene ya una larga tradición, como la fiabilidad y el control de calidad, a menudo surgen nuevos desafíos a la incorporación de la Estadística. Así, la degradación de productos en dinámica de medioambiente conduce a modelos como son los procesos puntuales multivariantes, para los cuales la inferencia está apenas desarrollada. Algunos investigadores comentan que hay una clara divergencia entre los métodos de control de calidad que los estadísticos proponen y los que realmente se necesitan, aunque en mi opinión, lo que en realidad se observa es una gran laguna entre el desarrollo metodológico y la investigación aplicada en esta área. Tukey ya comentaba que la investigación estadística no se debe concentrar en “buscar soluciones exactas a problemas erróneos sino en aproximar soluciones a problemas correctos”.

Uno de los desafíos a los que se enfrentan en la actualidad los estadísticos industriales está dedicado a la reducción del tiempo dedicado a estimar la vida de un producto. Los

investigadores están cambiando la experimentación física por construcción de modelos virtuales y técnicas CAD de diseño asistido en el ordenador. Esto representa un conjunto de oportunidades para los estadísticos, variando desde la verificación y validación de modelos al diseño eficiente de experimentos virtuales de alta dimensión. Por otra parte, los estadísticos han estado a menudo trabajando de modo aislado, y desarrollando enfoques parciales que ignoran información importante sobre procesos de manufactura multietapa secuencial. La complejidad de ambientes de fabricación avanzada implica que los estadísticos debemos trabajar junto con los ingenieros y tomar un enfoque de sistemas para mejorar el proceso. La combinación del análisis estadístico de los datos que entran en el proceso con la información sobre la arquitectura y secuencia conjunta, permite que la causa de la variabilidad del proceso sea diagnosticada más rápida y eficientemente.

Esta revisión que acabamos de hacer muestra las enormes oportunidades para los estadísticos de contribuir a realizar avances en las ciencias físicas y la ingeniería, mientras trabajan en la vanguardia de la metodología estadística. Hay una creciente sensación en este campo de que algunas ramas con un núcleo computacional, tales como la minería de datos, son las únicas técnicas estadísticas apropiadas para el entendimiento de datos complejos y masivos. Pero esto no es tanto una amenaza como una oportunidad para aprender nuevos modos de trabajar con datos y construir modelos. Una continua y quizá más seria amenaza, es la creencia de que la teoría estadística es dura, de nivel elevado y con un fin intelectual más valioso que el trabajo en aplicaciones estadísticas serias, lo que proyectaría una visión de que los estadísticos dedicados a realizar aplicaciones necesitan desarrollar teoría, pero los estadísticos dedicados a la construcción de teorías no necesitan realizar aplicaciones.

Por el contrario, todo lo que acabo de exponer pone de manifiesto que al desarrollar metodología para resolver problemas concretos en aplicaciones reales se consigue un fin también intelectualmente valioso, y es la mejor manera de asegurar que una teoría merece la pena.

No quisiera terminar sin agradecer a todos ustedes su atención, y al profesor Elipe, que hoy cesa en sus labores de Decano de nuestra Facultad de Ciencias, por haberme confiado el honor de impartir esta lección el día de nuestro patrono San Alberto.

Muchas gracias.

