

ANOTACIONES SOBRE LA IMPORTANCIA DE LA ESTADÍSTICA EN LAS INGENIERIAS INFORMÁTICAS

Angel Igelmo Ganzo
Universitat de les Illes Balears
e-mail:angel.igelmo@uib.es

RESUMEN: Esta Ponencia justifica la docencia de la asignatura de *Estadística* en los *Estudios de Ingeniería Informática*. La asignatura, *Estadística*, es troncal a nivel de Ingeniería Técnica pero desigualmente distribuidos, pues en Informática de Sistemas la carga docente es de 4,5 créditos mientras que en Informática de Gestión es de 9 créditos. Aunque es una asignatura consolidada según los vigentes planes de estudios es necesario insistir en la importancia de la misma justificando la docencia en las Ingenierias Informáticas recurriendo al propio valor de la disciplina como herramienta tecnológica. Por tanto en los apartados siguientes hacemos una referencia a la metodología estadística y a la metodología científica y técnica, haciendo ver que la Estadística es el Método Científico de las Ciencias de la Naturaleza y por tanto de la Ingeniería

1.- EL MÉTODO ESTADÍSTICO VERSUS EL MÉTODO CIENTÍFICO

1.1.- El Método estadístico

Sin más preámbulo asumimos lo escrito por Box, Hunter and Hunter (1) en la primera página del primer capítulo de su libro *Estadística para Investigadores*, cuando nos dicen que:

La investigación científica es un proceso de aprendizaje dirigido. El aprendizaje avanza con la interacción bidireccional entre hipótesis (conjeturas, modelo, teoría) y datos (hechos, fenómenos). Una hipótesis inicial conduce, por un proceso de deducción, a ciertas consecuencias que pueden ser contrastadas con datos. Cuando las consecuencias y los datos no concuerdan la discrepancia puede conducir, por un proceso inductivo a la modificación de la hipótesis. Se inicia entonces un segundo ciclo de iteración. Se deducen las consecuencias de la hipótesis modificada y se comparan de nuevo con los datos que a su vez pueden llevar a ciertas modificaciones y ganancia de conocimiento. Esta proceso de aprendizaje puede también describirse como un bucle de realimentación en el que la discrepancia entre los datos y las consecuencias de la hipótesis H_1 conduce a la hipótesis modificada H_2 , H_2 conduce a H_3 y así sucesivamente.

En base a lo anteriormente expuesto podemos definir las siguiente etapas en la aplicación de la metodología estadística, de acuerdo con Fernández-Abascal (2) y otros (10), (11):

i.- *Planteamiento del problema.* Se trata de definir las variables a observar, las cuales concretarán el problema, su ámbito, su alcance, etc. En definitiva es imprescindible saber lo que se va a estudiar. Un planteamiento inadecuado del problema nos puede llevar a hermosos errores.

ii.- *Obtención de datos.* La toma de datos es una fase importante del proceso pues una mala toma de datos puede invalidar todo el proceso, llegando a conclusiones erróneas. En las ciencias experimentales e ingenierías será una labor a desarrollar en laboratorio o in situ. Es evidente que en esta fase juega un papel clave el concepto de muestreo.

iii.- *Análisis descriptivo de datos.* Se trata de recoger, ordenar o clasificar y analizar los datos, función propia de la Estadística Descriptiva y del más moderno Análisis Exploratorio de Datos. que denominamos Análisis Descriptivo de Datos.

iv.- *Modelo teórico.* A la vista de las conclusiones obtenidas en el apartado anterior, se pasa a proponer un modelo teórico, recurriendo a técnicas de contraste de hipótesis, estimación de parámetros, pruebas no paramétricas, etc. Esta etapa es la que conocemos con el nombre de Estadística Inferencial (proceso inductivo).

v.- *Explotación de modelo.* Si no encontramos un modelo que se ajuste a los datos obtenidos, será conveniente volver a la fase i o ii, revisando tanto el planteamiento del problema como la obtención de datos. Si disponemos finalmente de un modelo que se ajuste a nuestra observación empírica procederemos a sacar conclusiones de todo orden a partir de nuestra modelización. Esta fase corresponde a un proceso matemático deductivo

Pero esta secuencia de etapas coincide sustancialmente con el método científico como describiremos más adelante. Por esta motivo algunos han definido que la Estadística es la **“Metodología de la Ciencia Experimental”**.

Abundando en esta forma de ver la Estadística, su metodología, también es interesante recurrir a L. Sachs (8) cuando nos dice:

Para nosotros es importante lo siguiente: basándonos en las condiciones específicas del problema hacemos suposiciones acerca del modelo subyacente y acerca del correspondiente modelo estadístico. Una vez comprobada la compatibilidad de las observaciones con el modelo estadístico se pasa a la descripción estadística de la población: En ambos casos se llega a conclusiones de tipo probabilístico, es decir, en términos de probabilidad. Así pues, la misión de la estadística consiste en encontrar los modelos estadísticos adecuados a cada problema y en aprovecharlos para extraer toda la información esencial que puedan encerrar los datos; es decir, la estadística proporciona modelos para la reducción de la información.

En todo este procesos o esquema hay, en mi opinión, una palabra clave: **DATOS**, de manera que parodiando a Arquímedes, podemos decir: **“Dadme datos y moveré al mundo”**.

Mencionemos lo que dice D. Peña (4) sobre la metodología Estadística:

La Estadística actúa como disciplina puente entre los modelos matemáticos y los fenómenos reales. Un modelo matemático es una abstracción simplificada de una realidad más compleja y siempre existirá cierta discrepancia entre lo observado y

lo previsto por el modelo. La Estadística proporciona una metodología para evaluar y juzgar estas discrepancias entre la realidad y la teoría.

Esta claro por tanto que el estudio de la Estadística es fundamental para todos aquellos que deseen trabajar en ciencia experimental, pura o aplicada (ingeniería), que requiera del análisis de datos y del diseño de experimentos, habiéndose dicho, incluso, que **la Estadística es la "tecnología" del método experimental.**

Estamos de acuerdo en que la Estadística es como una tecnología, aunque en todo momento usa un importante aparato matemático, por lo que algunos están tentados en presentar la estadística como una cosa matemática. Pero eso se denomina Estadística Matemática y no debe ser nuestro objetivo. Precisamente para que el alumno capte la capacidad tecnológica de la Estadística, debemos seguir un método didáctico en nuestras explicaciones diarias con los alumnos, que procure seguir la idea manifestada por Sixto Rios (7) en su Prologo de Métodos Estadísticos:

Creemos, por otra parte, que el papel de la Estadística Matemática, en relación con otros cursos de Matemáticas, puede ser importante para modificar la tendencia a cultivar en la enseñanza de la Matemática solamente las facetas algorítmicas y lógico-deductivas. La manera tradicional de enseñar las Matemáticas como un edificio acabado, partiendo de unos axiomas y demostrando una serie de teoremas para terminar cada lección con algunos ejercicios que van de lo ingenioso a lo trivial, pero que están tan lejos de la realidad y de las aplicaciones como la teoría escuetamente expuesta, es totalmente inadecuado y deformante para el que pretende trabajar como matemático aplicado, físico o ingeniero. El tipo tradicional de ejercicios de fin de capítulo, que suelen comenzar: "probar que...", "resolver la ecuación...", está muy distante de las preguntas que el matemático se encontrará en la realidad que serán del tipo "¿que sucede si...?", "¿Cuál es la relación que existe entre...?", "¿Cómo podemos encontrar...? Si queremos preparar a los alumnos para ser capaces de matematizar los problemas del mundo real, en que es frecuente que el objetivo central aparezca oscurecido por una masa de información incompleta y en que precisamente lo principal es destacar dicho objetivo, debemos cambiar radicalmente la manera tradicional de enseñar la Matemática. Del triple proceso de conceptualización, razonamiento lógico-deductivo y desconceptualización, que se presenta en toda aplicación de la Matemática al estudio de los fenómenos reales, se olvidan frecuentemente en la enseñanza la primera y última fases, que son justamente las más importantes,...

Por este motivo debemos hacer especial hincapié en los aspectos primero y tercero, siendo el aspecto matemático un asunto instrumental, pero sin que ello implique pérdida de rigor.

1.2.- La Estadística como metodología científica

El objeto de las *ciencias experimentales puras o aplicadas* no son características o sucesos aislados y únicos que afectan a un determinado individuo o elemento, sino experiencias repetibles. Los métodos estadísticos son necesarios en aquellas situaciones en que unos resultados no se pueden reproducir ilimitadamente y con toda exactitud, quedándonos en una situación de incertidumbre.

Otra forma de definir por tanto la Estadística es con relación a la toma de decisiones en un contexto de incertidumbre: “*La Estadística es un compendio de métodos que nos permite tomar decisiones óptimas en casos de incertidumbre*”. De aquí que la metodología estadística sea una herramienta imprescindible en la metodología científica y técnica.

Sin embargo es más habitual definir la estadística a partir de lo que hace, es decir como la *ciencia que recoge, ordena y analiza los datos de una muestra extraída de una población, para inducir conclusiones sobre el comportamiento de la población, valiéndose del cálculo de probabilidades*. Es decir, de las observaciones hechas sobre una parte de la población (muestra) se sacan conclusiones sobre la totalidad de la población; esto es, se procede inductivamente. Es fundamental que la muestra haya sido extraída al azar. Las muestras extraídas al azar, muestras aleatorias, son importantes, puesto que sólo de ellas se pueden extraer conclusiones acerca de la totalidad de la población. Hay diversos motivos por lo que no es posible analizar la totalidad de la población: por ejemplo, el análisis puede ser de tipo destructivo, o ser muy costoso o llevar mucho tiempo.

La definición dada de estadística permite distinguir dos etapas. La parte que recoge, ordena o clasifica y analiza los datos de la muestra, es lo que llamamos *Estadística Descriptiva*. Mientras que la parte que induce conclusiones sobre la población a partir del análisis de la muestra es lo que llamamos *Estadística Inferencial* (o Inferencia Estadística). Esta permite obtener regularidades generales cuya validez rebasa el conjunto de observaciones de la muestra.

En todas las ciencias empíricas la estadística permite la interpretación de datos empíricos y la verificación de teorías científicas o técnicas al confrontar resultados empíricos con resultados deducidos de modelos probabilísticos teóricos, considerando que todas las conclusiones posibles son de tipo probabilístico.

Abundando en estas ideas P.C. Mahalanobis dice que “*las leyes científicas no son formuladas por el principio de autoridad o justificadas por la fe o la filosofía medieval; la estadística es el único tribunal de apelación del conocimiento nuevo*”.

La ciencia trata con el conocimiento de los fenómenos naturales y su perfeccionamiento. Tal conocimiento se expresa en formas de leyes, axiomas o teorías, las cuales permiten la predicción de futuros acontecimientos. Buscamos hipótesis de trabajo, que estén apoyados por los hechos observados, los cuales en el transcurso del tiempo pueden ser reemplazados por mejores hipótesis, con más pruebas que las evidencien a partir de un grupo de datos más amplio. Según Popper (5) el método de investigación científico implica el siguiente ciclo sin fin:

$$P_1 \rightarrow CT \rightarrow EE \rightarrow P_2$$

donde P_1 y P_2 son la teoría inicial y su modificada respectivamente. CT representa el proceso de comprobación de la teoría y EE la eliminación de errores. La teoría inicial P_1 nace del fallo del conocimiento humano recibido anteriormente.

Este esquema permite aludir a la actitud de Popper ante la ciencia que lejos de considerarla como una ocupación rutinaria y prosaica, coloca a los científicos a la altura de los poetas y los músicos, por cuanto a imaginación creadora se refiere.

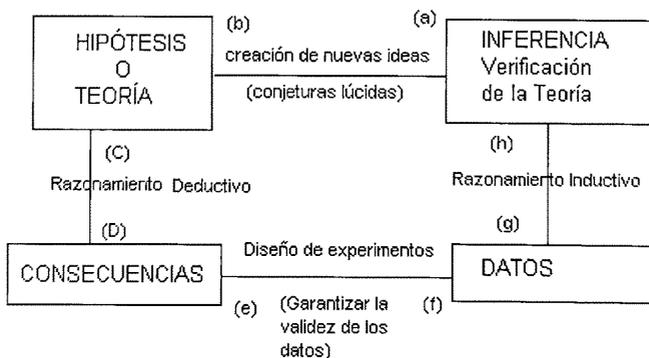
Pero incluso la metodología estadística trasciende a la ciencia, llegando a la filosofía. Así, Vargas Llosa (9) refiriéndose a Popper manifiesta:

Creo que no hay filosofía contemporánea o ningún pensamiento contemporáneo que haya hecho de la libertad un hecho tan absolutamente central como es el caso

de la filosofía de Popper. La idea de la verdad en Popper a mi me recuerda mucho un mito con el que se abre uno de los grandes libros de la antropología moderna, *La Rama Dorada* de Sir James Frazer. Recuerdo cómo se inaugura ese libro, describiendo en no se que cultura del Pacífico a ese reyezuelo que daba vueltas muy alarmado entorno al árbol, símbolo del poder en su cultura, esperando aquel que vendría a sucederlo, aquel que, como había hecho él con su antecesor, un día aparecería, lo mataría y lo sucedería y heredaría su poder. Leyendo a Popper, la idea de la verdad en Popper, yo he pensado siempre en ese reyezuelo. Porque, según Popper, la verdad tiene, como en el caso de este rey primitivo, ese carácter provisional: es una verdad que reina mientras no viene otra a liquidarla, a acabar con ella y a reemplazarla.

Esa verdad, que es una creación humana más que un descubrimiento, que es una hipótesis o teoría que pretende responder una pregunta y que entra luego a vivir una vida precaria, constantemente sometida a esa prueba del ensayo y del error, y que un día es incapaz de resistir esa presión y sucumbe a la prueba y es derrotada y reemplazada por otra, es una verdad que hace de la crítica, es decir, el derecho a someter a todas las verdades establecidas al fuego graneado, constante, continuo de la refutación, la condición básica del conocimiento. Ninguna verdad puede ser reemplazada sin la prueba del ensayo y del error, y no puede haber prueba del ensayo y del error sin la crítica, es decir, sin la libertad, sin el derecho a refutar todas las verdades establecidas sin excepción.

El método científico, siguiendo a Rao (6), se puede representar gráficamente como sigue:



donde se pone de manifiesto los dos procesos lógicos: razonamiento deductivo y razonamiento inductivo. En este diagrama se ponen de manifiesto dos etapas: por un lado los pasos (a) → (b) y (c) → (d), donde están involucrados cada área de investigación, y por otro lado (e) → (f) y (g) → (h) que están sujetos a la metodología estadística.

A través de la recolección de datos relevantes y válidos, la estadística permite al científico poner en juego todas sus potencialidades creativas y así descubrir nuevos fenómenos. Los métodos estadísticos han sido de gran valor especialmente en aquellas ciencias y técnicas donde el rango de variación de las observaciones es a menudo grande y el número de observaciones suele ser limitado. Resumiendo: **“Si no existiese la Estadística, habría que inventarla”**.

2.- LA INFLUENCIA ACTUAL DE LA ESTADÍSTICA

Desde principios de siglo un sinnúmero de nuevos métodos y teorías y un inmenso campo de aplicación dificultan una visión general de lo acontecido y hace arriesgado resaltar alguna innovación decisiva. Sin embargo repasemos algunos acontecimientos.

Hasta la segunda guerra mundial se extiende la aplicación de la estadística en áreas tan diversas como la Ingeniería (Control de Calidad por Shewart, métodos de predicción y control de procesos y codificación de señales por Wiener y Shanon), la Física (teoría cinética de los gases), la Antropología, la Psicología o la Medicina. La búsqueda de respuesta a los nuevos interrogantes planteados por estas aplicaciones impulsan, a su vez, el desarrollo de nuevos métodos estadísticos. Los problemas agronómicos llevan a Fisher a crear la teoría de diseños experimentales y un problema de discriminación en antropología (estudio de cráneos) le lleva a inventar el análisis discriminante. El análisis factorial surge ligado a problemas en Psicología mientras que las ciencias sociales impulsan el desarrollo de métodos para analizar muchas variables simultáneamente (análisis multivariante). Los problemas en Ingeniería conducen a un estudio sistemático de la teoría de modelos dinámicos (procesos estocásticos), y a la teoría de predicción y de extracción de señales de Wiener y Komogorov en la década de los cuarenta. Las necesidades de control de procesos sugiere a Pearson la creación de la teoría general de contraste de hipótesis, junto con Neyman. El trabajo del Statistical Research Group en Princeton en aplicaciones industriales de la estadística durante la segunda guerra mundial condujo a Wald a inventar los contrastes secuenciales para el control de recepción y fue básico en el nacimiento y desarrollo de la teoría estadística de decisión. Durante este periodo las aplicaciones de la estadística a la economía conducen a una disciplina con contenido propio: la Econometría.

A partir de la segunda guerra mundial se puede decir que empieza la época moderna de la Estadística, apareciendo aspectos diferenciadores respecto de la época anterior, de los que cabe resaltar (ver (10), (11)):

- El espectacular desarrollo de las matemáticas, que proporciona nuevos recursos formales a la Estadística.
- Los nuevos campos de aplicación, principalmente la Biología, Economía, Industria, etc. que abren nuevas perspectivas a la Estadística.
- La generalización masiva del uso de los ordenadores y de la informática que introduce la Estadística en problemas prácticos hasta ahora inabordables por imposibilidad de cálculo, a la vez que impulsa teorías apropiadas para el análisis de grandes colecciones de datos.

Esta breve revisión de la evolución de la Estadística muestra como los grandes periodos de avances teóricos se han producido generalmente ligados a la resolución de importantes problemas prácticos. Paradójicamente, la mayoría de las contribuciones importantes que hemos revisado son debidas a investigadores que no pueden calificarse exclusivamente como estadísticos, sino en el sentido mas amplio del término como científicos.

3.- BIBLIOGRAFÍA

1. Box, G.E.P., Hunter, W.G., Hunter, J.S. (1989). Estadística para investigadores. Introducción al diseño de experimentos, análisis de datos y construcción de modelos. Editorial Reverte.
2. Fernández-Abascal, H., Guijarro, M.M., Rojo, J.L., Sanz, J.A. (1994) Cálculo de Probabilidades y Estadística. Editorial Ariel.
3. Igelmo, A. (1998) Análisis Descriptivo de Datos. Servicio de Publicaciones de la Universidad de las Islas Baleares.
4. Peña Sánchez de Rivera, D. (1994). Estadística, Modelos y Métodos. Tomo I: Fundamentos. Alianza Universidad. 2ª Edición.
5. Popper, K.R. (1973). La lógica de la Investigación Científica. Editorial Tecnos. 3ª reimpresión.
6. Rao, C.R. (1994). Estadística y verdad. PPU.
7. Rios, S. (1973). Métodos Estadísticos. Ediciones del Castillo.
8. Sachs, L. (1978). Estadística Aplicada. Editorial Labor.
9. Schwartz, P., Rodríguez Braun, C. y Méndez Ibisate, F. (1993). Encuentro con Karl Popper. Alianza Editorial
10. A. Meca, M. Almiñana, F. Borrás, J. Cánovas, N. Llorca, E. Molina, D. Morales, A. Sánchez, J.V. Segura. (1998). Sobre el futuro de la Estadística y la Investigación Operativa: Primera Parte. Boletín de la SEIO. Volumen 14, nº 3, septiembre de 1998, pag. 2-9.
11. A. Meca, M. Almiñana, F. Borrás, J. Cánovas, N. Llorca, E. Molina, D. Morales, A. Sánchez, J.V. Segura. (1998). Sobre el futuro de la Estadística y la Investigación Operativa: Segunda Parte. Boletín de la SEIO. Volumen 14, nº 3, septiembre de 1998, pag. 2-9.