

© 2004, Ing. Miguel Ángel Gallardo Ortiz en [WWW.CITA.ES](http://WWW.CITA.ES)  
Apartado (P.O. Box) 17083, E-28080 Madrid, Spain  
Tel: 914743809 Fax: 914738197 Móvil: 619776475

## **ESTADÍSTICA CON ECONOMETRÍA Y PROSPECTIVA PARA EL PRÍNCIPE** “Dígasele con datos (sin equivocarse demasiado)”

### **Introducción general e interés de la Estadística en PCs**

Si Nicolás Maquiavelo estuviera actualizando ahora mismo “EL PRÍNCIPE”, muy probablemente dedicaría dos capítulos, uno para las verdades útiles, y otro para las falsedades (más útiles aún), de la Estadística del Estado, instruyendo sobre todo lo que, estadísticamente, debe de saber un príncipe. Sabemos que nuestro príncipe heredero, Felipe de Borbón y Grecia, ha recibido una educación muy completa en Estadística y especialmente en su aplicación más sensible políticamente, la Econometría, para la más acertada toma de decisiones en prospectiva política, económica, social e internacional.

El Príncipe de Asturias aprendió temas relacionados con la estadística primero con el Profesor Dr. Enrique Fuentes Quintana, luego en la Universidad Autónoma (aunque tuvo muchas asignaturas de Derecho, se complementaron con varias de Economía) y posteriormente también en la Georgetown University de Washington DC donde estudió relaciones internacionales que exigen números y estimadores estadísticos cada vez más precisos). Puede asegurarse que la Estadística es, hoy por hoy, la asignatura clave fundamental para gobernantes y jefes de Estado, en todo el mundo. Por eso creemos que a los príncipes (a todos los príncipes, y a ese pequeño príncipe que todo profesor, jefe, cliente o juez lleva dentro), hay que “decírselo con datos (sin equivocarse demasiado)”.

Para el internauta, la Estadística parece ser la asignatura pendiente, y no porque falten buenas fuentes de información y excelentes programas, sino porque nunca hay el tiempo y el sosiego que la Estadística merece. También echamos en falta buenos tutoriales que nos permitan distinguir, estadísticamente, el grano de la paja, las luces de las sombras, las voces de los ecos, y los datos de estimaciones u opiniones, más o menos fundadas.

En la Georgetown University (¡sí, sí, precisamente en la misma universidad en la que estudió nuestro Príncipe!) hay un encantador profesor, John C. Pezzullo, Ph.D, que se ha dedicado principalmente a la investigación en química y bioestadística, y ahora que se ha jubilado, ha elaborado, con gran generosidad científica, una utilísima recopilación de referencias que pueden verse en <http://members.aol.com/johnp71/javastat.html>

Ha sido un gran placer tratar por correo electrónico con este profesor de Georgetown, que nos ha animado y orientado para poner a la disposición de nuestros lectores algunas referencias estadísticas dignas de un príncipe. Pero antes de empacharnos de datos y de instalar sofisticados programas informáticos, hay que estudiar algunos fundamentos matemáticos que aquí trataremos de hacer tan entretenidos y útiles como sea posible.

Los príncipes, y los que no lo somos, necesitamos cada vez más la estadística, que los matemáticos reconocen que tuvo su propio príncipe: Carl Friedrich Gauss (1777-1855).

Gauss fue un niño prodigio que pronto hubiera podido ocupar una cátedra universitaria, pero prefirió dedicar su vida profesional a la dirección del Observatorio Astronómico de

Göttingen más por aversión a la enseñanza, que por auténtica afición astronómica. Dicen sus numerosos biógrafos que sólo durmió una vez fuera de su casa durante los últimos 27 de sus 78 años de vida, y que disfrutaba enseñando sólo a los que mostraban auténtico interés, pero rehuía cualquier tipo de reconocimiento o relevancia social. Pero sus contribuciones al álgebra, teoría de números, y a lo que aquí nos ocupa son trascendentales. Hay un antes y un después de Gauss en la historia de las probabilidades que no es sino la lucha de la racionalidad contra los dioses (recomendamos aquí el fascinante libro “Against the gods: The remarkable story of risk”, de Peter L. Bernstein para quienes quieran conocer la apasionante historia de la estadística).

### **Fundamentos generales de Estadística**

Aristóteles, que entre otras muchas cosas, fue el educador de Alejandro Magno, consideraba que las tres formas ideales de conocer, y de dar a conocer, se basaban en definiciones, divisiones y argumentaciones con el mayor rigor lógico posible. Estamos seguros de que los libros de texto de Matemáticas del lector tienen algunas buenas definiciones, divisiones o índices de materias y lecciones de Estadística, y nos gustaría que en este momento surgiera en la memoria del lector las frases que utilizó el primer profesor que tuvimos para interesarnos por la Estadística, o para examinarnos de ella, que no es lo mismo. Aquí no hay examen, pero sí estudio.

El diccionario de la Real Academia Española define Estadística (de estadista), así:

1. f. Censo o recuento de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de un Estado, provincia, pueblo, clase, etc.
2. Estudio de los hechos morales o físicos que se prestan a numeración o recuento y a comparación de las cifras a ellos referentes.
3. Mat. Ciencia que utiliza conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades.

Si precisamos esta definición general triple inspirándonos con la más específica de la Enciclopedia de las Matemáticas de la Academia de Ciencias de la URSS (en la espléndida edición en castellano de Antonio Rubiños), podemos convenir en que la Estadística Matemática trata de los métodos matemáticos de sistematización, elaboración y uso de los datos estadísticos para INFERIR (repetimos: INFERIR) conclusiones científicas y prácticas. Al igual que entre Álgebra y Cálculo, es muy importante distinguir entre las inferencias estadísticas, y las deducciones matemáticas, porque las inferencias que no son inducciones matemáticas perfectas, excepcionalmente pueden sorprendernos (aunque parezca que es la realidad la que sorprende, pero lo cierto es que a la realidad siempre le da absolutamente igual si hemos inferido correctamente o no, porque la Ciencia, cualquiera que sea la especialidad de la Ciencia, nunca se equivoca, sino que somos nosotros los que nos equivocamos, porque somos los únicos que podemos equivocarnos, aunque “hagamos como que nos sorprende el habernos equivocado”), mientras que las deducciones y las inducciones matemáticas, cuando son correctas, no admiten ninguna sorpresa, lógicamente.

Así, en la ciencia de lo probable, las inferencias Estadísticas proporcionan “metadatos” (datos sobre los datos), pero como ocurre con cualquier otro tratamiento de la información, su precisión siempre dependerá de la que tengan los datos de partida. Por lo tanto, el valor de toda inferencia estadística depende de los datos utilizados, y

también del método aplicado para su tratamiento, que en este siglo XXI, está en un PC, o impreso en papel, salvo en muy contadas excepciones.

Y el trabajo estadístico no termina ahí. La buena educación en Estadística también se nota cuando leemos, y sobre todo, cuando interpretamos los resultados estadísticos, porque no basta con calcular automáticamente aplicando fórmulas matemáticas, sino que es muy necesario, y cada vez más, el ser capaz de analizar desde múltiples perspectivas, cualquier inferencia estadística que sirva de base para una decisión importante. El arte de interpretar se llama “hermenéutica”, y resulta útil tanto para hacer un comentario de texto, como para analizar inteligentemente cualquier tipo de estadística. Y este arte hermenéutico nos permite diferenciar bien las estadísticas inteligentes, de las que no lo son, y también las honestas, de las perversas.

Otras importantes diferencias deben establecerse entre “estimables” encuestas de opinión o de intención, y rigurosas estadísticas retrospectivas, porque los sondeos estadísticos, como bien saben los sociólogos, están sometidos y expuestos al criterio y credibilidad de los encuestados, y por lo tanto, siempre ofrecen incertidumbres, mientras que las estadísticas retrospectivas pueden realizarse con el máximo rigor y exhaustividad descriptiva informáticamente, de manera que si los datos son precisos, y están completos, el cálculo estadístico produce resultados exactos. Afortunadamente, la informática personal ahora hace mucho más fácil el cálculo estadístico, que hasta hace pocos años, podía requerir auténticos superordenadores al alcance de muy pocos. Sin embargo, el análisis exige muy complejos conceptos y técnicas de inteligencia artificial.

Es, precisamente, el exceso de capacidad y de potencia informática lo que explica que muchas estadísticas muy necesarias no se hagan, o se cometan tremendos errores, o que se contaminen con tópicos o fuentes poco fiables, o en definitiva, que la estadística disponible oficialmente no sea todo lo inteligente que desearíamos. Cuando se pueden hacer tantas cosas, acaba por no hacerse ninguna, y cuando hay tantos medios para hacerlas bien y pronto, resulta que se hacen muy mal, y muy tarde, o no se hace nunca.

Pero empecemos por el principio, dando a los datos de partida la máxima importancia.

### **Adquisición, formateo, calidad, interpretación y presentación de datos estadísticos, en Excel**

Existen muy diversas fuentes de datos públicos útiles para realizar cálculos estadísticos, aunque lamentablemente suelen estar poco actualizadas y sólo sirven para hacer análisis histórico, pero en este apartado nos centraremos en el trabajo de conversión de formatos, sea cual fuere su fuente. La práctica hace que se reduzcan tiempos, esfuerzos, y sobre todo, temibles errores de conversión, porque las opciones de ciertos programas de uso muy habitual para el trabajo estadístico acaban resultando familiares y seguras. Por ejemplo, Excel ofrece una opción de “DATOS” y dentro de ella “Obtener datos externos” para hacer (o repetir) consultas a páginas de Internet, o bien importar a la hoja electrónica activa datos procedentes de ficheros de múltiples formatos, con distintos separadores o diversos criterios de selección para su consideración y los cálculos posteriores.

El método más sencillo es el manual, o bien “cortar“ (Ctrl+C) y “pegar” (Ctrl+V) por columnas o filas de datos marcados, y el más complejo, posiblemente, sea ajustar una

consulta, con reconocimiento inteligente de la fecha de actualización, de una página de Internet con mucha información compleja, interpretada, calculada y presentada casi automáticamente. Cuando se consigue esto último, la satisfacción que proporciona ver trabajar telemáticamente a Excel “importando datos” puede resultar comparable a la que experimenta un maestro cuando su alumno, príncipe o no, hace bien las cosas. De alguna manera, tenemos que “enseñar a Excel” a entenderse con los datos, empezando por cortar y pegar, y acabando por configurar consultas eficaces, en tiempo real, en Internet, mediante un modelo adecuado a nuestras fuentes y propósitos, que acaba concretándose en una consulta directa a la fuente (mientras rezamos para que nuestra fuente no cambie por sorpresa su formato y nos deje “fuera de juego”).

Por supuesto, Excel también puede acceder a bases de datos (MS-Access, pero también Dbase, FoxPro, etc) y tal vez haya que pasar los datos de la aplicación (por ejemplo, una contabilidad, o aplicación SQL del Server de Microsoft, ORACLE, Software AG, mainframes de IBM, Unix u otros sistemas) a un formato texto con separadores, pero la ventaja de realizar la presentación en Excel está en la facilidad con la que se pueden ilustrar los resultados mediante gráficos de tipos estándar o personalizados. Además, una vez que tenemos los resultados en Excel, resulta muy fácil llevarlos a PowerPoint para realizar presentaciones auténticamente espectaculares, aunque también tenemos experiencia en lo difícil que resulta recorrer el camino inverso, porque lo que nos da una presentación en PowerPoint no suele ser suficiente para componer una hoja Excel con la que podamos reconstruir un modelo estadístico con los mismos resultados.

Es decir, que lo fácil es pasar un trabajo estadístico de Excel a PowerPoint, pero es mucho más difícil lo contrario (que es lo que algún principesco jefe nos pide a veces mostrándonos lo que le han dado en una presentación y diciéndonos, simplemente “hay que hacer esto mismo, y mejor, para mañana”), porque por lo general faltan los datos de partida y las fórmulas aplicadas, y tratar de generar un modelo que se ajuste a lo que se nos presenta en PowerPoint exige un trabajo de análisis y de diseño de una hoja Excel.

La calidad de los datos es un aspecto esencial en estadística. Por lo general, el resultado es tan malo como lo sea el peor de los datos, pero siempre resulta muy difícil detectar un error en los datos. Sólo la experiencia, el sentido común y la lógica analítica, pueden ayudarnos a detectar y filtrar eficazmente alguna celda, fila o columna de datos sospechosos, pero suprimir un dato es una forma de manipular la estadística. La detección y el tratamiento de los errores y sus contaminaciones presenta serios dilemas a la deontología estadística, porque el error intencionado es una especie de terror estadístico, y más aún cuando la solución que se le da a lo que parece un error es demasiado interesada, porque algunas correcciones son peores aún que los errores.

Si se trabaja en Excel, o con bases de datos con volúmenes de información manejables y la complejidad del modelo estadístico lo permite, lo más recomendable es mantener siempre una versión de todos los datos en crudo, con todos sus errores, y hacer un informe explicativo de todos los filtrados que se hagan, comentando con precisión para futuras necesidades. Para el analista objetivo, no existe su propio criterio, sino que hay varios criterios aplicables a cada conjunto de datos que presentan anomalías, y por lo tanto, lo más profesional es dar una versión de los resultados para cada criterio de selección aplicado.

Excel resulta muy recomendable, por lo tanto, para realizar Estadística descriptiva, e incluso tiene una opción para hacer de una sola vez, un resumen de estadísticas generando un campo en la tabla de resultados por cada una de las siguientes variables estadísticas: Media, Error típico (de la media), Mediana, Moda, Desviación estándar, Varianza, Curtosis, Coeficiente de asimetría, Rango, Mínimo, Máximo, Suma, Cuenta, Mayor (#), Menor (#) y Nivel de confianza.

Pero Excel ofrece también sofisticadas herramientas de análisis estadístico (que hay que instalar cuando no aparecen en el menú de herramientas) como complemento o macro automática para realizar operaciones muy complejas.

Para dominar todas las herramientas estadísticas de Excel con sus respectivas opciones habría que realizar un curso completo de Estadística, pero para ilustrar la utilidad de alguna de estas herramientas vamos a destacarla con ejemplos actuales, que se verán más adelante como regresiones lineales.

Sobre Excel se han elaborado plantillas para representación de histogramas, cálculos y técnicas analíticas muy variadas que pueden ser completamente gratuitas (freeware) o bien versiones de evaluación (shareware) o para uso no comercial con restricciones que limitan su funcionalidad respecto a la versión comercial. Una interesante colección de cada una puede verse en <http://members.aol.com/johnp71/javasta2.html#Excel>

### **La revolución estadística descriptiva de PC-Axis**

Los países escandinavos (Suecia, Noruega, Dinamarca y Finlandia), junto al Canadá, lideran las innovaciones estadísticas, por lo que muchos países europeos adoptan metodología y herramientas proporcionadas por los nórdicos. Un ejemplo de ello es el programa descriptivo PC-Axis y sus aplicaciones especiales, geográficas de PX-Map y para Internet de PX-Web, desarrolladas en Suecia por “Statistics Sweden”, que es el organismo oficial de la estadística nacional de Suecia.

El software se puede descargar desde Suecia en <http://www.pc-axis.scb.se>  
Y también en el Web del Instituto Nacional de Estadística [www.ine.es](http://www.ine.es)

Sin embargo, la versión que se distribuye gratuitamente no incluye las opciones y las funciones para la creación de archivos, por lo que el PC-Axis básico recuerda al Adobe Acrobat Reader para uso estadístico, de manera que es relativamente fácil leer y convertir a Excel los datos, pero no podemos elaborar fácilmente tablas. PC-Axis puede incorporar software complementario, como por ejemplo Formula One y First Impression de Visual Components, pero para el usuario familiarizado con Excel lo más recomendable es convertir la tabla elaborada con PC-Axis a un archivo.xls o bien a un formato algo más orientado a la estadística como es el NSD-Stat o el GESMES (EDIFACT) para el intercambio electrónico de datos.

No solamente el INE está utilizando oficialmente PC-Axis en España. El Consejo General del Poder Judicial elabora un CD-ROM con la memoria anual de las Estadísticas Judiciales con PC-Axis que también pueden consultarse (aunque con muchas más dificultades) en Internet y es previsible que este estándar de software aumente su ámbito de aplicación para visualizaciones de datos en estadística descriptiva, en varias instituciones más que realizan publicaciones electrónicas. Pero

repetimos que entre el usuario y el creador de PC-Axis hay diferencias muy importantes, y necesidad de más funciones que no se proporcionan gratuitamente.

PC-Axis presenta estructuras de tablas estadísticas con las variables y valores que intervienen en cada una, y lo hace con el modo de presentación y funciones típicas de una hoja de cálculo, aunque con mucha menos flexibilidad que Excel, porque está especialmente orientada a la estadística, y no a la ofimática más general. Pueden configurarse diversos aspectos de visualización e impresión de los datos y se aprovechan las posibilidades que ofrece el entorno Windows. PC-Axis usa el portapapeles, permite seleccionar aplicaciones ofimáticas instaladas en el sistema usuario para traspasarles en cualquier momento los datos que se están presentando (Excel, Word, Bloc de Notas, etc) y exporta cualquier tabla o vista a múltiples formatos, entre ellos los de texto plano y Excel.

La estadística descriptiva frecuentemente acaba en la infografía de las cifras. PC-Axis es comparable, al menos en sus funciones puramente descriptivas, a Excel, y también muy compatible, en posibilidades gráficas e infográficas para quien sabe lo que quiere evidenciar, aunque no es comparable a Excel en su integración ofimática con Word o PowerPoint, de manera que resulta recomendable exportar a formato Excel las tablas y trabajar en él las ilustraciones y las presentaciones PowerPoint.

En las ediciones en CD-ROM o en las que se descargan de páginas Web puede accederse a las tablas de datos, bien desde el programa PC-Axis, abriendo la correspondiente base de datos, o bien desde el icono de cada publicación, que da acceso a la información sobre las características de la investigación estadística, metodología, cuestionarios e índice de tablas. Como en otros muchos programas, entre el talento, la tenacidad, el método y la práctica se encuentra el éxito, pero en el caso de PC-Axis, el acceso a recursos técnicos especiales, y a la información fresca es fundamental para cocinar bien los datos estadísticos (no es lo mismo ser el primero en disponer de la memoria del CGPJ del año 2004, que tener que trabajar ahora sólo con datos de 1996).

### **Conceptos esenciales de estimadores, y regresiones lineales**

La estimación estadística es la función de variables aleatorias que se emplea para evaluar los parámetros desconocidos de una distribución teórica de probabilidades. Realmente lo que se calcula es la formulación del error que se trata de minimizar, presuponiendo que los errores son puramente aleatorios, y siguen una distribución normal de probabilidades descubierta por Gauss. La curva normal es imprescindible para cualquier cálculo actuarial, demográfico, sociológico, biológico o minero, y es la base para realizar estimaciones, por ejemplo, de tendencias lineales.

Es un peligroso error confiar a ciegas en los resultados de las estimaciones estadísticas porque como decía un gran maestro, no se debe abordar el cálculo de ningún problema cuya solución aproximada no se conozca de antemano. Hay que saber lo que se está calculando y en qué rango de valores pueden estar los resultados, porque un mínimo error informático, sólo detectable por el sentido común y la lógica de las proporciones que da la experiencia, puede ocasionar gravísimas consecuencias y perjuicios. Por ello, hay que dar dos miradas a los libros de matemáticas y tres a la realidad, por cada vistazo al manual del programa informático que utilicemos para hacer estimaciones estadísticas.

Básicamente, una regresión lineal obtiene la recta que mejor se ajusta a una nube de puntos por el método de los “mínimos cuadrados”, de manera que la suma de las distancias de todos los puntos a esa recta sea la menor posible, suponiendo que los errores y su medida son aleatorios. Esta recta, que resulta característica de cada distribución, tendrá mayor o menor utilidad dependiendo de lo más o menos alineados que estén los puntos, siendo máxima cuando todos los puntos estén exactamente en esa recta (sin ningún error), y mínima, o nada relevante en absoluto, cuando los puntos estén, por ejemplo, en los vértices de un triángulo equilátero, cuadrado, pentágono o con simetría radial en una circunferencia (cuando el error relativo es máximo), porque en estos casos, no hay recta que pueda representar la distribución correctamente. La distancia de los puntos reales a los más próximos a la recta viene a representar el error, por lo que minimizando la suma de los cuadrados de tales distancias a la recta que se trata de obtener lo que se hace es minimizar el error que se comete al sustituir la realidad por la estimación que proporciona la recta ajustada, y su error característico.

Un concepto muy importante es el nivel de confianza asociado a la regresión lineal, que por lo general se estima en el 95% (y es éste el valor que utiliza Excel por defecto en su herramienta para realizar sus regresiones lineales). Sin embargo, el estudio de ciertos riesgos exige mayor seguridad, y conviene tener presente qué nivel de confianza se está utilizando en cada regresión lineal, para ser consciente de cómo ha de interpretarse bien.

Evidentemente, la utilidad de las regresiones lineales en las ciencias experimentales es enorme, y de hecho, la eficacia de los fármacos más especializados se evidencia mediante la distinta correlación de distintos indicadores entre pacientes que reciben un tratamiento comparándolos con los grupos de control que no lo reciben. Es decir, que para comprobar si una regresión lineal entre posibles causas y posibles efectos, conviene tener en cuenta, inteligentemente, los casos más extremos de ausencia de causa, y de ausencia de efecto, así como los casos de una sola presencia, además de la cuantificación más precisa posible de ambas presencias respecto a las de control.

En estos momentos, en España se está debatiendo, a veces sin la suficiente seriedad y sin rigor estadístico, sobre riesgos más o menos correlacionados con ciertos datos que pueden ser muy preocupantes.

El 25 de noviembre de 2003 la presidenta del Consejo de Seguridad Nuclear (CSN), María Teresa Estevan Bolea, anunció, en su comparecencia anual ante la Comisión de Economía del Congreso de los Diputados, que se realizará un estudio epidemiológico en 500 municipios con el fin de comprobar la incidencia de las centrales nucleares sobre la salud de las personas y comparar los datos de cáncer en las poblaciones próximas a esas instalaciones con los de otras localidades, según informa la agencia Europa Press y puede comprobarse en las actas parlamentarias.

Varios investigadores de las Ciencias de la Salud ya han hecho estudios de muy diversas enfermedades, y en particular sobre el cáncer, considerando algunos indicadores y posibles factores cancerígenos, pero el gran problema es el de la disponibilidad de datos precisos que permitan correlacionar la aparición de cáncer con sus causas. La inmensa mayoría de los certificados de defunción indican que la causa de la muerte es el “paro cardíaco” (así el médico de turno nunca se equivoca), pero la

“causa de la causa” suele ser otra, y las regresiones son la herramienta estadística más precisa para indagar “causas de causas” en investigaciones científicas.

Incluso cuando se dispone de muchos datos relevantes, hay un problema de parametrización medico-sanitaria, porque las escalas de las variables pueden ser correlacionadas de muchas maneras, siendo unas más correctas que otras. Por lo tanto, la dependencia de una variable puede ser directamente lineal, o bien proporcionalmente al cuadrado, o al logaritmo, o a la raíz cuadrada, o a una función trigonométrica, o a cualquier operación matemática realizada con la otra. La historia de la medicina, y la de su estadística, tiene una importante referencia en la famosa biblioteca de aplicaciones FORTRAN BMDP (BioMeDical Programs) cuya utilidad calidad, desde hace décadas, ha impulsado numerosos estudios científicos, y sus fórmulas matemáticas siguen estando en muchos programas desarrollados en otros lenguajes más modernos.

En el caso del estudio sobre los riesgos de cáncer, nos resulta evidente que no puede haber una dependencia clara, y menos aún de tipo lineal, con la distancia a las centrales nucleares, porque intervienen otros muchos factores fenomenológicos más, como el tiempo de exposición, edad, otras actividades, etc. Además, la leucemia, o el cáncer de próstata (que sólo sufren los hombres), o el de mama (que sólo sufren las mujeres), pueden mostrar muy diversas dependencias de distintos factores cancerígenos, por lo que nos encontramos ante una explosión combinatoria de muy numerosas hipótesis de dependencia de los diversos tipos de cáncer, respecto de un inimaginable conjunto de posibles causas de cada cáncer, y sus interrelaciones combinadas, que a su vez pueden representarse, en distintas escalas, mediante muchísimos indicadores paramétricos, por lo que cualquier hipótesis razonable, necesariamente, ha de ser “multivariante”.

Las regresiones multivariantes, especialmente las no lineales, por lo general, son mucho más complicadas que las que se ajustan bien a una recta, que Excel sí puede hacer inmediatamente, incluso calculando y representando valores futuros mediante los datos disponibles en sólo dos rangos (columnas). Cuando un efecto depende de dos causas X e Y, también puede depender de las posibles correlaciones entre esas dos causas (función por lo tanto de X e Y) y el modelo es complicada hasta el punto de requerir programas altamente especializados en el análisis multivariante, en aplicaciones informáticas que a veces son capaces de encontrar los límites operativos del sistema y de capacidad de almacenamiento y ejecución de programas en los ordenadores más potentes. Y cuando los ordenadores aumentan su potencia, como los estadísticos nunca van a estar plenamente satisfechos con los recursos técnicos disponibles, ya se están planteando nuevas necesidades en el cálculo numérico, que actualmente tiene como primer límite los 32 bits de los procesadores más comunes, con los que, en principio, “sólo” pueden direccionarse 4 Gbytes de memoria.

Pero las ideas estadísticas esenciales son las mismas, tanto si trabajamos en un modesto PC con Excel, como si utilizamos un superordenador, y sólo quien es capaz de hacer, y explicar, correctamente una regresión lineal simple puede pretender que se comprendan los casos multivariantes más complejos, porque la mejor informática no sustituye al estudio de los fundamentos matemáticos, aunque facilite y acelere los cálculos.

La estadística sin inteligencia es casi tan peligrosa como la inteligencia sin estadística, y no bastan las matemáticas para hacer buenos planteamientos estadísticos. El concepto fundamental para el investigador que trata de correlacionar rigurosamente unos



indicadores con otros para hacer estimaciones y pronósticos científicos dependen, antes que de sus conocimientos matemáticos o informáticos, de su visión fenomenológica, porque es el fenómeno lo que se pretende modelizar estadísticamente en una regresión, analizando retrospectivamente los datos disponibles, para hacer predicciones que, si se cumplen, confirman nuestras hipótesis, y hacen más creíbles nuestros futuros pronósticos, para poder tomar medidas correctoras que reduzcan los riesgos (por ejemplo, de contraer una enfermedad) y aumenten los beneficios (de la salud). Hay que leer mucho, y deliberar, y calcular, y probar, y pensar, sobre todo, pensar, para plantear correctamente la regresión, antes de empezar a equivocarnos, con o sin informática.

Nosotros esperamos y deseamos que el estudio sobre cáncer y radiaciones anunciado por el Consejo de Seguridad Nuclear (CSN) sea “fenomenal”, porque si hay algo en lo que la estadística debe ser fenomenal, es en lo que garantiza la salud de todos, queridos príncipes.

### **Series temporales y estudio de tendencias cíclicas**

Algunos fenómenos dependen del tiempo con tendencias periódicas. Para analizar los datos periódicos hay que estudiar cuidadosamente sus posibles frecuencias.

Una serie temporal es una sucesión de valores observados de una variable en momentos del tiempo diferentes, que pueden ser ordenados cronológicamente.

Las series temporales resultan apasionantes para el estudio estadístico, especialmente cuando se dispone de suficiente información histórica, homogénea y actualizada. Pero hay que ser muy cuidadoso para detectar cambios de índices de referencia (por ejemplo, las pesetas no han valido lo mismo hace 3 años, que hace 25, y por eso se convertían a “pesetas constantes”), cambios metodológicos (por ejemplo, la Encuesta de Población Activa ha considerado de manera distinta a lo largo de la historia, a los parados de larga duración, a los que voluntariamente se salen del mercado de trabajo, o a quienes están buscando activamente un trabajo que no encuentran, y con ciertas pautas cíclicas, a quienes sólo trabajan algunos meses del año, y en consecuencia, los subsidios por desempleo se han administrado de maneras muy diversas en la historia reciente), y también la existencia de lagunas o datos perdidos que deben ser estimados por interpolación, o por técnicas más complejas de “arqueología de la información” (por ejemplo, aproximando las estimaciones de los datos que faltan consultando hemerotecas o bibliotecas, o preguntando a posibles testigos que nos ilustren sobre los períodos de tiempo en los que faltan datos).

En definitiva, la experiencia demuestra que muy rara vez puede aplicarse directamente un programa de análisis de series temporales sin un cuidadoso trabajo previo sobre los datos, sus frecuencias más o menos evidentes, y todas las interpretaciones que empezaron cuando el hombre empezó a preguntarse por el movimiento de los astros, con muy pocos datos entonces, hasta que Tycho Brahe proporcionó a Kepler, en unas tablas astronómicas, las primeras “crónicas marcianas” con las que se formularon las tres leyes que siguen explicando con gran precisión el movimiento de todos los astros del Universo en órbitas elípticas. Probablemente fue Kepler quien hizo la primera inferencia de la estadística matemática sobre un conjunto de datos amplio, iniciando una carrera en la que tomaron su testigo Newton con sus leyes, después Gauss con su

clarividencia estadística y Einstein con su teoría de la relatividad hasta hacer de la Física Estadística una exigente disciplina en el estado del arte de la Ciencia y la Técnica.

Hoy sigue siendo mucho más importante obtener sencillas leyes que explican bien complejas relaciones entre numerosos sujetos y objetos, que listados incomprensibles e inacabables, por lo que hay que aprender a inferir estadísticamente para encontrar sentido a muchas de las funciones para el análisis de series temporales que ofrecen los programas especializados, entre los que puede destacarse EVIEWS.

Aunque intuitivamente podemos encontrar fenómenos cíclicos (muchos indicadores de actividad son claramente estacionales permitiendo distinguir fácilmente entre comportamientos de verano e invierno, otros son mensuales, y otros semanales porque ciertas cosas que ocurren los miércoles no ocurren los domingos, por ejemplo), y Excel ofrece algunas opciones como las transformadas de Fourier, además de gran flexibilidad en el manejo y presentación de datos, cualquier estudio serio de series temporales necesita una herramienta especializada.

Eviews es un programa estadístico diseñado especialmente para el estudio de series temporales. El objeto básico de trabajo en Eviews es el “workfile” (archivo de trabajo). Los workfiles tienen dos características esenciales: FRECUENCIA y RANGO. Las series temporales suelen ser observadas en intervalos de tiempo de la misma longitud, o frecuencias, en el calendario. Cuando se especifica una frecuencia, se le comunica al programa el intervalo temporal entre observaciones. Eviews maneja datos anuales, semestrales, trimestrales, mensuales, semanales y diarios (con semanas de 5 ó 7 días), o irregulares (series de sección cruzada), en las que hay que analizar separando las distintas frecuencias que se interfieren.

Eviews destaca por su facilidad para hacer histogramas representativos ilustrando especialmente simetrías y curtosis (grado de alargamiento de una curva de distribución) pero Eviews hace mucho más que estadística descriptiva para series temporales, porque está diseñado para comprobar multitud de hipótesis mediante test aplicables al workfile.

El contraste de las hipótesis ajusta la expresión de la serie temporal a un modelo, pero conviene tener claros algunos conceptos que diferencian las series naturales (por ejemplo, las astronómicas o meteorológicas), de las econométricas (económicas).

### **Modelos Econométricos**

Las inferencias sobre numerosos casos bien analizados estadísticamente conducen a leyes en Física, Biología, Sociología o Economía. No hay muchas leyes generalmente aceptadas y la historia de la Economía demuestra que ciertos errores pretendidamente científicos pueden tener efectos desastrosos. El primero de todos consiste en no diferenciar Economía de Econometría, sin recordar que la Ciencia (también la Ciencia Económica) nunca se equivoca, sino que nos equivocamos nosotros, siempre.

El modelo econométrico exige especificaciones estadísticas mucho más precisas que los económicos, más basados en valores y tendencias, que en series de datos. Por lo tanto, los modelos económicos resultan ser demasiado generales para ser aplicados a momentos, lugares y actividades económicas concretas, que es lo que se pretende con un modelo econométrico. Los economistas y los econometristas han creado una sopa de

letras para modelos empresariales, sectoriales, regionales, nacionales y supranacionales (ARIMA, VAR, VEC, ARCH o nombres propios como Klein, Wharthon, Theil, Makridakis o Wheelwright y por supuesto, los equilibrios de Nash) denominan sofisticadas técnicas y programas informáticos en modelos econométricos).

Pero hay una idea muy importante, imprescindible, para enfocar correctamente el estudio econométrico, y es la de la relatividad de las leyes, porque a diferencia de lo que puede observarse en Astronomía o Meteorología, en la seguridad de que siempre obedecerá a las mismas leyes, las leyes económicas y los modelos econométricos están muy condicionadas por las circunstancias en el momento en el que se aplican las supuestas leyes, hasta el punto de que ciertos modelos sólo tienen sentido en lugares muy concretos mientras están vigentes ciertas normas. Como decía el poeta, “el ojo que tú ves no es ojo sólo porque tú lo veas, sino que es ojo porque te mira”, y cuando los agentes económicos detectan que la situación responde a un modelo econométrico, rápidamente reaccionan haciendo que ese modelo ya no tenga la misma validez.

Así, la estadística para la econometría encuentra ciertos límites para su modelización que sólo pueden ser superados mediante la teoría de juegos que supone inteligencia y conocimiento a todos los agentes. John Nash, premio Nobel de Economía en 1994, y más famoso por la película “Una mente maravillosa”, revolucionó la teoría de juegos con sus modelos (los llamados “equilibrios de Nash”), en los que todos los agentes pueden considerar lo que otros por su parte harán para tomar sus propias decisiones.

El error económico puede ser más desastroso aún cuando los estadísticos se obstinan en seguir utilizando un modelo obsoleto y en ocasiones la informática mal aplicada es la responsable de muchas aberraciones económicas. Una vez que los feroces operadores económicos conocen las claves de un modelo, especulan sobre él, y si es fácilmente simulable la especulación por ordenador, quebrarán cualquier estrategia en su beneficio.

No es recomendable adquirir un costoso paquete de software altamente especializado en econometría para tratar de utilizar alguno de los modelos que ofrece, sino que antes al contrario, lo que se debe de hacer es estudiar profundamente la fenomenología, antecedentes y referencias técnicas para elegir el modelo econométrico adecuado. Después, y sólo después, deben buscarse los programas que mejor satisfacen las necesidades técnicas precisadas por el modelo, porque también sería un tremendo error buscar una princesa por los requerimientos del trono de la futura en lugar de intentar que el trono sea el que se adapte a nuestra princesa, para que, al menos, le sea llevadero cuando llegue a ser reina. Además, las cosas, y más aún las personas, acaban siendo como tienen que ser, y no de otra manera, ¿verdad, querido príncipe?

### **Prospectiva y futuro para los príncipes**

Más Allá de los objetivos económicos, el Estado tiene todas las obligaciones que la Constitución intentó garantizar al ciudadano, que resultan completamente imposibles no sólo de satisfacer, sino ni siquiera de plantear, si no existe una buena estadística oficial.

Lamentablemente, cuando la opinión pública no se fía de las estadísticas (y suele haber buenas razones para ello), resulta mucho más difícil la administración de cualquier recurso, porque sin indicadores objetivos no puede planificarse nada correctamente.

Por eso, el futuro es siempre sorprendente. Y no podía serlo menos el futuro de la futurología, o la prospectiva de la prospectiva. Lo que sí podemos asegurar, es que la informática personal está revolucionando a la inteligencia oficial para la que trabaja una parte de la estadística oficial. Actualmente hay sistemas expertos para decidir qué modelo econométrico es el más adecuado para una determinada predicción. Es decir, que hay programas (como por ejemplo FORECAST PRO, MASTER, FOCA, FORMAN, SIMPC, MICRODYNAMO, etc) que escogen a su vez a otros programas para estudiar las diferentes probabilidades de casi todos los futuros posibles. El problema de la predicción se convierte así en predecir cuál será el mejor método de predecir, y para eso hay una competencia salvaje entre los asesores de todo príncipe.

El futuro de la futurología, y la prospectiva de la prospectiva, están en algún PC, porque el espíritu del mago Merlín y de todos los adivinos que los príncipes han tenido a lo largo de la Historia, se encuentra ahora representado por modelos basados en intensas estadísticas que caben en un disco no demasiado grande. Esperamos y deseamos que quien tenga ese PC futurólogo tenga también acceso directo al correo electrónico del Príncipe de Asturias, o al de quien pueda decírselo con datos, y a tiempo.

Y es que, como decía un ex-presidente del Gobierno de España a un periodista y académico de la lengua llamado Juan Luis Cebrián, en un libro titulado “EL FUTURO YA NO ES LO QUE ERA”, pues eso, que el futuro ya no será lo que creíamos que iba a ser, como seguro que Vd. ya sabe muy bien, querido príncipe.

Miguel Ángel Gallardo Ortiz, E-mail: [miguel@cita.es](mailto:miguel@cita.es)  
Diplomado en Altos Estudios Internacionales,  
Ingeniero de Minas y Criminólogo en [www.cita.es](http://www.cita.es)

## CUADROS PROPUESTOS

### 1. Fórmulas matemáticas de una regresión lineal simple

La ecuación de la línea de regresión de un modelo simple es:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Donde:

$y$  = La variable que va a ser modelada. Se conoce como la “variable dependiente” o “response variable” o “variable de respuesta” o “efecto”.

$x$  = Es la variable predictora. Se conoce también como la variable independiente, o “causa”.

$\varepsilon$  = Epsilon. Es el componente del error aleatorio (el malo al que hay que eliminar).

$\beta_0$  = Beta Cero. Es el punto en el cual la línea intercepta el eje de ( $y$ ).

$\beta_1$  = Beta Uno. Es la pendiente de la línea. Es la cantidad por la que cambia el componente determinístico de ( $y$ ) por cada unidad que ( $x$ ) incremente.

En estas condiciones, la regresión lineal se obtiene calculando:

$$\begin{aligned} \beta_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{(\sum x_i y_i - n \bar{x} \bar{y})}{\sum x^2 - n (\bar{x})^2} = \\ &= \frac{\sum x_i y_i - [(\sum x_i)(\sum y_i) / n]}{[\sum x^2 - (\sum x)^2 / n]} \end{aligned}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = (\sum y_i / n) - [\beta_1 - (\sum x_i / n)]$$

## **2. Herramientas estadísticas en MS-Excel para**

- Realizar un análisis de varianza
- Correlación
- Covarianza
- Suavización exponencial
- Análisis de Fourier
- Prueba t para varianzas de dos muestras
- Realizar un análisis de la prueba t
- Histograma
- Media móvil
- Generación de números aleatorios
- Jerarquía y percentil
- Regresión
- Muestreo
- Prueba z para medias de dos muestras

### **3. Libros recomendados (todos con CD-ROM)**

Problemas de Inferencia Estadística.  
Estadística Empresarial con Microsoft Excel  
De Isabel Parra Frutos, Editorial AC, 2ª Edición, 2003

Estadística para Administración  
Berenson, Levine y Krehbiel, Ed. Prentice Hall, 2ª Edición, 2001

Modelos Econométricos  
A. Pulido y J. Pérez, Ed. Pirámide, 2001

Predicción y simulación aplicada a la economía y gestión de empresas  
A. Pulido, Ed. Pirámide, 1999

## **4. Referencias y Metarreferencias en Internet**

Software Estadístico Gratuito (FREEWARE)  
enlazado desde <http://www.cita.es/estacita>

Paquetes de propósito estadístico general

WinIDAMS desarrollado por UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) para el análisis estadístico disponible en <http://www.unesco.org/webworld/idams/>  
Los particulares pueden solicitarlo en [http://www.unesco.org/webworld/portal/idams/individual\\_request.html](http://www.unesco.org/webworld/portal/idams/individual_request.html)  
y las instituciones en [http://www.unesco.org/webworld/portal/idams/institution\\_request.html](http://www.unesco.org/webworld/portal/idams/institution_request.html)

La última versión de WinIDAMS de UNESCO es la 1.1 y está disponible en español, inglés y francés.  
(hay problemas para conseguirlo, porque NO LO ENVÍAN solicitándolo según indican)

Instat Plus -- Desarrollado en University of Reading del Reino Unido en <http://www.ssc.rdg.ac.uk/software/instat.html>

OpenStat -- Win 95/98/NT, estilo SPSS-like con interface de usuario y excelente manual en PDF  
<http://www.statpages.org/miller/openstat/>  
<http://www.statpages.org/miller/openstat/OS2pdfManual.zip#B>

EasyStat -- Para Windows y Mac con t-tests, F-tests, simple ANOVA, tablas de contingencia, modelos, Mantel-Haenszel.  
Disponible en <http://www.geocities.com/Tokyo/Ginza/1276/>

ViSta -- Visual Statistics para Win3.1, Win 95/NT, Mac y Unix, con Structured Desktop y Análisis Estadístico en  
<http://forrest.psych.unc.edu/research/>

AM -- Buen análisis de muestras complejas de gran tamaño con estimación MML y aproximación de series de Taylor  
<http://am.air.org/>

Para GNU/LINUX existen muy diversas herramientas. En la pág. 178 del N° 158 de PC Actual (diciembre 2003) VNU LABS describió un apartado a la Estadística GNU/LINUX en el artículo "Software Científico" mencionando PSPP y RLAB, así:

PSPP -- Alternativa libre al software comercial SPSS en <http://www.gnu.org/software/pspp/>

RLAB -- Implementación libre del lenguaje S desarrollado en los Laboratorios Bell por John Chambers. Aunque nació como un lenguaje de programación orientado a cálculos estadísticos, nos permite modelización tanto lineal como no lineal, análisis de series temporales, contrastes de hipótesis, etc. Además, permite enlazar código escrito en C, C++ o Fortran para su utilización en tiempo de ejecución. RLAB está en <http://rlab.sourceforge.net>

Complementos para MS-Excel

PopTools (muy completo), en <http://www.dwe.csiro.au/vbc/poptools/>

OZGRID contiene más de 4.000 hojas electrónicas en <http://ozgrid.com/>  
Para descargas gratuitas <http://www.ozgrid.com/download/default.htm>

DE Histograms en [http://www.aecouncil.com/data\\_analysis/](http://www.aecouncil.com/data_analysis/)

Exact para intervalos de confianzas en muestras de distribución Binomial y Poisson en <http://members.aol.com/johnp71/confint.xls>

Essential Regression en <http://www.geocities.com/SiliconValley/Network/1900/>

ASCA Profile Analysis en <http://espse.ed.psu.edu/spsy/Watkins/Software/asca.sea.hqx>

Correlation Tests en <http://espse.ed.psu.edu/spsy/Watkins/Software/Testcorr.sea.hqx>

Metarreferencias (referencias de referencias)

Free Statistical Software (actualizado a 01/06/2004) en <http://members.aol.com/johnp71/javasta2.html>

Econometric Software en enlaces del "Econometrics Journal" en <http://www.feweb.vu.nl/econometriclinks/software.html>  
Lista de distribuidores de software en <http://www.feweb.vu.nl/econometriclinks/software.html#Distributors>

DERECHO ESTADÍSTICO y  
ESTADÍSTICA FORENSE  
en <http://www.cita.es/apuntes>

Para más información y actualizaciones de <http://www.cita.es/estacita>  
se agradecerán cuantos mensajes se envíen a [miguel@cita.es](mailto:miguel@cita.es)  
Tel.: 914743809 y móvil 619776475



## **Relación de imágenes de pantallas**

Símbolos.bmp

Guía de símbolos del libro Problemas de Inferencia Estadística.  
Estadística Empresarial con Microsoft Excel  
De Isabel Parra Frutos, Editorial AC, 2ª Edición, 2003  
(también enviado en Excel para poder editar mejor)

Business Statistics.bmp

README.doc del CD-ROM (está en inglés aunque el libro esté traducido)  
Estadística para Administración  
Berenson, Levine y Krehbiel, Ed. Prentice Hall, 2ª Edición, 2001

EViews.bmp

Guía EViews del CD del libro  
Modelos Económicos  
A. Pulido y J. Pérez, Ed. Pirámide, 2001

Estados Financieros.bmp

Estados Financieros Provisionales de una Empresa Periodística  
Del CD del libro  
Predicción y simulación aplicada a la economía y gestión de empresas  
A. Pulido y Ana Mª López, Ed. Pirámide, 1999

Free Statistical Software.bmp

en <http://members.aol.com/johnp71/javasta2.html>

UNESCO.bmp

WinIDAMS desarrollado por UNESCO (Organización de las Naciones Unidas para la  
Educación, la Ciencia y la Cultura), en <http://www.unesco.org/webworld/idams/>

PC-AXIS en INE.bmp

PC-AXIS, el formato de las publicaciones electrónicas  
Para descargar PC-Axis en <http://www.ine.es/prodyser/pcaxis/pcaxis.htm>

ECONOMETRICS.BMP

Econometric Software en enlaces del "Econometrics Journal" en  
<http://www.feweb.vu.nl/econometriclinks/software.html>

ESTACITA.bmp

<http://www.cita.es/estacita>  
Los principales enlaces referenciados en este artículo

© 2004, Ing. Miguel Ángel Gallardo Ortiz en [WWW.CITA.ES](http://WWW.CITA.ES)  
Apartado (P.O. Box) 17083, E-28080 Madrid, Spain  
Tel: 914743809 Fax: 914738197 Móvil: 619776475