

Datos

Boletín de la Asociación Mexicana de Estadística

Número 26, Año 12, Abril 2003



A M E

En este número:

- ✓ **Experiencias de Posgrado**
Universidad de Minnesota
página 2
- ✓ **Noticias**
página 4
- ✓ **Galería de los Grandes**
página 7
- ✓ **Navegando**
página 8
- ✓ **Humor**
página 8

Editor:

Eduardo Castaño Tostado
ecastano@uaq.mx

Co-editores:

Karim Anaya Izquierdo
karim@sigma.iimas.unam.mx

Manuel Mendoza Ramírez
mendoza@itam.mx

Asociación Mexicana de Estadística
IIMAS-UNAM

Depto. de Probabilidad y Estadística
Apartado Postal 20-726 Admon. 20
Del. Álvaro Obregón
CP 01000 México D.F.

amestad@amestad.org.mx

<http://amestad.org.mx>

Editorial

Eduardo Castaño

¿Hacia dónde?

La situación de la investigación en México es incierta debido a los cambios acelerados que como sociedad mundial y local estamos enfrentando; en el estado actual de cosas se puede palpar inconformidad en la comunidad científica nacional que ha señalado en general a las políticas en materia de ciencia y tecnología nacionales como erróneas o poco claras; en contraparte se cuestiona a la investigación de gran parte de la comunidad científica por su resistencia a considerar criterios de pertinencia local ante las grandes carencias nacionales. Mucha de la situación actual de la dirección y la calidad de la investigación tienen que ver con cuestiones del financiamiento. Algunas cifras reveladas en reportes de la OECD pueden dar claves en el caso de México. En el año 2002 México ocupaba el último lugar entre los países pertenecientes a la OECD en el gasto en investigación y desarrollo como porcentaje del producto interno bruto con apenas el 0.5 % (a pesar que el congreso apruebe en papel el 1 %). La mediana en la OECD es casi 2 %. En ese año, por abajo del 1 % sólo se encuentran la Grecia olímpica, Turquía, Polonia y República de Eslovenia; destaca Suecia con más de 4 % dedicado a la investigación y desarrollo. En cuanto al origen del financiamiento se muestra que en México, en el 2002, la iniciativa pri-

vada financia casi el 30 % de la investigación y desarrollo y el 60 % se financia por el gobierno. Sólo Grecia, Hungría, Polonia y Portugal mostraron porcentajes similares. Los demás miembros del OECD muestran porcentajes de financiamiento privados por arriba del 37 %. El financiamiento otorgado por el gobierno mexicano para investigación, sin embargo, y de acuerdo a los resultados de las convocatorias a concurso por fondos de los últimos años, está restringiendo el acceso a fondos en dos vertientes. La primera vertiente lleva a la reducción de fondos para investigación básica (cifras de la Academia Mexicana de Ciencias muestran que 16 % de proyectos sometidos son ahora financiados, cuando el histórico era 40 %); así sólo grupos muy consolidados obtienen fondos, ya que ante la restricción de fondos para investigación básica, la propia comunidad científica ha endurecido los criterios en el sentido de sólo avalar proyectos de investigación que sean originados por investigadores nacionales de nivel dos hacia arriba; Darwinismo científico. La segunda vertiente de la política de financiamiento por el gobierno mexicano, es condicionar el acceso a fondos si se tienen usuarios potenciales de la investigación. Se topa con sectores privado y gubernamentales con una perspectiva incipiente de la necesidad de investigación y por ende con pocos proyectos de investigación. En cuanto al número de patentes por miles de habitantes, la cifra de México ni siquiera es reportada por la OECD, suponemos por ser prácticamente cero. La mediana es aproximadamente 0.025, siendo el líder Suecia con un valor ligeramente superior a 0.1. En lo referente a la gente, en el rubro del número de investigadores por cada mil empleados, México vuelve a ocupar el último lugar en la OECD apenas con un investigador por cada 1000 empleados; el máximo lo muestra Finlandia con 16, y la mediana es aproximadamente seis. ¿Cómo atraemos a jóvenes a hacer carrera ante las cifras descritas? Si alguien cuenta con otras cifras que muestren una visión diferente de la investigación en México, realmente serían muy bienvenidas en la búsqueda de balances.

U

¹Jorge de la Vega Góngora es Matemático egresado de la UNAM. Ha desarrollado sus actividades profesionales fundamentalmente en el área de operaciones de banca central. Obtuvo el Doctorado en Estadística (Ph. D.) en la Universidad de Minnesota en el año 2000 y actualmente ocupa la posición de investigador de mercados en el Banco de México.

Una experiencia de posgrado en la Universidad de Minnesota

Jorge De la Vega¹
jvega@banxico.org.mx

En junio de 1996, recibí mi carta de aceptación al programa de doctorado en la escuela de estadística de la Universidad de Minnesota, en el *campus* Twin Cities. Este ha sido uno de los días de mayor ilusión en mi vida. Una de las razones por las que solicité admisión en esta universidad fue porque reconocí algunos nombres de sus profesores en mis libros de texto de estadística en la Facultad de Ciencias de la UNAM, que me parecieron claros y muy bien escritos. Entre estos profesores se encuentran Bernard Lindgren, Luke Tierney, Dennis Cook y Sandford Weisberg, quien finalmente se convirtió en mi asesor. La Universidad de Minnesota es una de las 10 universidades públicas más grandes de los Estados Unidos, otras en su categoría incluyen a las Universidades de Pensilvania y Wisconsin. Aunque la universidad es grande, la escuela de estadística cuenta con una planta de cerca de 20 profesores y 50 estudiantes graduados de diversas nacionalidades, aunque predominan los estudiantes chinos. Una situación curiosa es que hasta la fecha, soy el primer y único estudiante mexicano de estadística en esta escuela. Esto no fue un problema para mí, ya que hay varios mexicanos y latinos que trabajan en otras áreas del conocimiento y con los que pude compartir mi experiencia y conocimiento para apoyarlos en sus propias áreas de investigación. Esta fue una parte muy enriquecedora de mi experiencia académica. El *campus* de la Universidad de Minnesota en las ciudades gemelas -Minneapolis y Saint Paul- en realidad está conformado por dos sedes comunicadas por un autobús que lleva de una a otra en cinco o 10 minutos, dependiendo del clima. El clima es un aspecto muy importante en el oeste medio de los EU, al grado de que la universidad ha tenido que suspender clases

por el frío o los peligros de tornado. Un compañero me dijo la primera vez que me vio llegar a una clase con mi chamarra de plumas de ganso a las siete de la mañana: ¡welcome to Minnesnowta! Otra broma muy común es que en Minnesota sólo hay dos estaciones: invierno y la de reparación de caminos. La escuela de Estadística tiene presencia en ambas sedes, ya que en Minneapolis se encuentra la sección teórica del departamento, fuertemente vinculada con la escuela de Matemáticas y con la escuela de Salud Pública. En Saint Paul, se concentra la sección de estadística aplicada que da soporte a los departamentos de biología, agronomía, economía aplicada, recursos naturales, etcétera. Como estudiante, hay que tomar clases en ambas secciones en forma cotidiana. De hecho, esto marca un aspecto importante del programa doctoral en estadística, que se encuentra fuertemente balanceado entre la teoría y la práctica. Algunas de las clases, como análisis de supervivencia y métodos empíricos Bayesianos se pueden tomar en el Departamento de Bioestadística. Algunos de los temas de investigación de los académicos son: diseño experimental, teoría de decisiones, modelos lineales generalizados, métodos no paramétricos, juegos y procesos estocásticos, muestreo Bayesiano. En el departamento también se organizan seminarios con mucha frecuencia y se invita a profesores de otras universidades y departamentos para enriquecer las discusiones y los puntos de vista. Las relaciones entre profesores y estudiantes son abiertas; prácticamente no encontré a ningún profesor que no participara, de alguna forma, en las convivencias con los estudiantes y por lo menos una vez al año se lleva a cabo una fiesta con todas las personas que pertenecen al departamento en la casa de alguno de los profesores. Aunque hay una tendencia a competir fuertemente entre los estudiantes, me fue posible discutir con mis compañeros diferentes formas de atacar problemas. La vida cotidiana en Minneapolis puede ser difícil en el invierno, cuando muchos estudiantes enfrentan un sentimiento de frustración debido a lo severo del clima y su larga duración que ocupa la mitad del año. Yo encontré que estas condiciones constituían un ambiente propicio para el estudio, ya que no hay muchas tentaciones para salir. Para mi fue muy agradable y relajante estudiar sentado cerca de una ventana viendo nevar o llover. Aún en el frío, se puede esquiar o patinar en los lagos. Sin embargo, en el verano se puede

disfrutar la ciudad pues tiene muchas facilidades para excursionar, usar bicicleta y nadar en alguno de los muchos lagos que se encuentran en la ciudad. La universidad cuenta con dos conjuntos habitacionales que rentan departamentos a estudiantes que cubren ciertos requisitos. Estos lugares son muy agradables para vivir y fomentan la colaboración y la vida comunitaria. La vida cultural es muy rica y accesible a los estudiantes. Más información sobre el programa doctoral, junto con la experiencia de otros estudiantes se puede encontrar en www.stat.umn.edu.



Mensaje del Presidente de la Asociación Mexicana de Estadística 2003 - 2005

Jorge H. Sierra Cavazos

Es para mi un honor fungir como presidente de la Asociación Mexicana de Estadística (AME). Es una responsabilidad muy grande, además de implicar mucho trabajo y esfuerzo. No es tarea fácil e intentaré dar lo mejor de mi parte para apoyar el trabajo que la AME ha estado haciendo durante todos estos años. Siempre he criticado a los políticos que hablan sólo por hablar y que prometen y nunca cumplen. No quiero ahora tomar ese papel y prefiero presentarme como uno más de los miembros de nuestra Asociación que tiene deseos de trabajar. Recientemente me llamó la atención un hecho que me sorprendió sobremanera. La clausura de un programa de posgrado en Estadística en la ciudad de Saltillo. No conozco las razones específicas que fueron dadas en este caso pero creo que no hay justificación aceptable para una acción de tal naturaleza. Es preocupante para nosotros como "comunidad estadística" del país, el hecho de que la cultura estadística esté cada vez más devaluada en México, al grado de llegar a extremos como éste. Tal parece que los estadísticos tuviéramos que resignarnos a ser una especie en extinción en nuestro país. Esto a pesar de los tremendos errores que se cometen constantemente en la recolección, análisis e

interpretación de datos en la práctica de la Estadística y que en muchos casos se muestran demasiado evidentes. Creo que como estadísticos en México no tenemos otra opción más que trabajar juntos. Unirnos como grupo colegiado para fortalecernos. Debemos hacer un esfuerzo y dejar de lado rencores personales o de grupo y debemos trabajar en unidad. Los exhorto a que trabajemos juntos, sin importar la unidad profesional en donde laboremos, apoyémonos unos a otros como estadísticos y tengamos para nuestros colegas críticas constructivas y palabras de aliento. No seamos nosotros mismos los causantes de nuestra propia destrucción.



Noticias

1. Extendemos una sincera felicitación al Dr. Manuel Mendoza Ramírez como nuevo miembro de la Academia Mexicana de Ciencias. La ceremonia se llevó a cabo en las instalaciones de la Academia el pasado 28 de abril de 2004.
2. Se recomienda la lectura del No. 5659 Vol. 303 del 6 de febrero de 2004 de la revista *Science*, que trae una sección especial sobre la relación entre la matemática y la biología, con artículos y ligas a la *web* interesantes.



Un ejemplo en Ox

Ramsés H. Mena Chávez²

En el número anterior de DATOS se introdujo, de manera breve, el lenguaje de programación matricial Ox creado por Doornik (2001). Este número tiene como objetivo ilustrar, a través de un ejemplo sencillo, cómo hacer una implementación en Ox.

Supóngase que se tiene una muestra de tamaño 1000 de una variable aleatoria cuya distribución admite una función de densidad dada por

$$f_X(x | \lambda) = 2\lambda K_0(2\sqrt{x\lambda}) I(x > 0), \quad (1)$$

donde $K_0(\cdot)$ denota una función modificada de Bessel de orden 0 y $\lambda > 0$. Dada esta muestra se quiere obtener el estimador máximo verosímil (EMV) para el parámetro λ .

De esta manera las funciones a implementar se pueden resumir en

1. Generar una muestra de tamaño 1000 de una variable aleatoria con distribución cuya densidad está dada por la expresión (1)
2. Dada la muestra, calcular el EMV para λ .

Primero se dará el código requerido para las dos tareas arriba mencionadas y después se dará una explicación.

```

1  /*****
2  /*****      Ejemplo.ox      *****/
3  /*****
4
5  /*****  Librerías requeridas  *****/
6  #include <oxstd.h>
7  #include <quadpack.h>
8  #include <oxdraw.h>

```

²Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM. A.P. 20-726. 01000 México D.F. Email: ramses@sigma.iimas.unam.mx

```

9  #import <maximize>
10 #include <../src/solver/ox>
11 /***** Declaraci\{o}n de funciones *****/
12 fX(const x, const lambda);
13 fXapp(const x);
14 fX(const x, const lambda);
15 fXapp(const avF, const x0);
16 SimX(const N, const lambda);
17 LogVS(const lambda, const x);
18 LogVSapp(const lambda, const AdFunc,
19 const avScore, const amHess);
20 /***** Variables globales *****/
21 decl Glb_lambda, Glb_U, Glb_X;
22 /***** Programa principal *****/
23 main()
24 {
25   decl Muestra, r, emv;
26   ranseed(2);
27   Muestra=SimX(1000, 1);
28   r=range(0,5,0.05);
29
30   DrawAdjust(ADJ_AREA_X,0,0,5);
31   DrawDensity(0,Muestra,"Ejemplo",0,1, 0, 0, 0, 999,2);
32   DrawMatrix(0, fX(r,1),{"$f_X(x\mid \lambda=1)$"},
33   r, "",0,1);
34
35   MV(0.05, &emv, Muestra);
36   println("estimador para lambda= ",emv);
37
38   DrawMatrix(0, fX(r,emv),{"$f_X(x\mid \hat{\lambda})$"},
39   r, "",0,5);
40   SetDraw(SET_LEGEND,1,1);
41   SaveDrawWindow("Ejemplo.eps");
42   SaveDrawWindow("Ejemplo.gwg");
43   ShowDrawWindow();
44 }
45 /***** Funcion de densidad *****/
46 fX(const x, const lambda)
47 {return 2*lambda*bessel(2*sqrt(x*lambda), 'K', 0);}
48
49 fXapp(const x)
50 {return fX(x,Glb_lambda);}
51 /***** Funcion de distribucion *****/
52 FX(const x, const lambda)
53 {
54   decl res, abserr;
55   Glb_lambda=lambda;
56   QAG(fXapp,0.001,x,2,&res,&abserr);
57   return res;
58 }
59
60 fXapp(const avF, const X0)
61 {avF[0] = FX(X0[0],Glb_lambda)-Glb_U; return 1;}
62 /***** Simula v.a. de X *****/
63 SimX(const N, const lambda)
64 {
65   decl U, x0, RS, SU, SUinx, i;
66   U=ranu(N,1); RS=zeros(N,1);
67   SU=sortc(U); SUinx=sortcindex(U);
68   x0=<0.0005>;
69   Glb_lambda=lambda;
70   for(i=0;i<=N-1;i++)
71   {
72     Glb_U=SU[i];
73     SolveNLE(fXapp, &x0);
74     RS[SUinx[i]]=x0;
75   }
76   return RS;
77 }
78 /***** Calcula log-verosimilitud *****/
79 LogVS(const lambda, const x)
80 {return sumc(log(fX(x,lambda)));}
81
82 LogVSapp(const lambda, const AdFunc, const avScore,
83 const amHess)
84 {AdFunc[0]=LogVS(lambda,Glb_X); return 1;}
85 /***** Estimador maximo verosimil *****/
86 MV(const lambda_0, const Pe, const x)
87 {
88   decl lw, P;
89   Glb_X=x;
90   P=lambda_0;
91   MaxControl(-1,100); MaxControlEps(1e-5,1e-5);
92   MaxBFGS(LogVSapp, &P, &lw, 0, 1);
93
94   Pe[0]=P;
95   return 1;
96 }

```

Para generar números aleatorios correspondientes a la densidad dada en (1) se utilizará el método de inversion, es decir, dado un número aleatorio U proveniente de una distribución uniforme en el intervalo $[0, 1]$, se necesita encontrar x tal que $U = F_X(x)$. Con esta finalidad se declaran las funciones `fX(const x, const lambda)` y `FX(const x, const lambda)` situadas en las líneas 44 y 50 del código. Para la función de densidad se utilizó la función `bessel(<argumento>, <tipo>, <orden>)`, que pertenece a la librería estándar `oxstd.h` de Ox y calcula la función de Bessel requerida en (1). Nótese que esta última función se puede aplicar a un argumento matricial, cualidad que se usa en la función que calcula el logaritmo de la verosimilitud (línea 77).

Aunque en este caso la función de distribución acumulada $F_X(x)$ puede tomar forma explícita en términos de otras funciones de Bessel, aquí se omite tal información con la finalidad de ejemplificar la fácil implementación de la función de integración numérica `QAG(...)` (línea 54), que a su vez se habilita con la inclusión de la librería `quadpack.h` en línea 7. Esta función, tal y como se encuentra dada en la línea 7, calcula la siguiente expresión

$$F_X(x) = \int_{0,001}^x f_X(t) dt. \quad (2)$$

El número 2 indicado en el cuarto argumento de `QAG(...)` denota la opción de 10-21 puntos en el método de Gauss-Kronrod, que es el utilizado por `QAG(...)`, el quinto y sexto argumento direccionan el resultado y el error absoluto del integrador a las variables correspondientes (en este caso `res` y `abserr` respectivamente). Una explicación de los argumentos en `QAG(...)` así como también otros métodos de integración se pueden consultar en la ayuda de Ox, situada típicamente en el archivo `..\Ox\doc\index.html`. También se puede consultar Piessens (1983) para una descripción más detallada de las rutinas utilizadas.

Con el objetivo de resolver $U = F_X(x)$ y maximizar la log-verosimilitud se han utilizado dos rutinas

numéricas. La primera `SolveNLE(...)`, que resuelve un sistema de n ecuaciones $f(x)$ no lineales con n incógnitas (en este caso $n = 1$, línea 71) se utiliza para resolver $F_X(x) - U = 0$ (ver función en línea 58). `SolveNLE(...)` requiere la inclusión de `../src/solvenle.ox` lo cual lleva a cabo en la línea 10.

Como rutina de maximización numérica la librería `maximize` (incluida en la línea 9) contiene el algoritmo de Broyden-Fletcher-Goldfarb-Shanno (BFGS), implementado mediante la función `MaxBFGS(...)` (línea 89). Información sobre este algoritmo se puede obtener en Press *et al.* (1992). Tanto `QAG(...)`, `SolveNLE(...)` y `MaxBFGS(...)` (líneas 54, 71 y 89 respectivamente) requieren como primer argumento las funciones en `Ox` proporcionadas de manera especial (líneas 47, 58 y 80 respectivamente). Nótese que para el algoritmo `MaxBFGS(...)` también se pueden proporcionar "Score" y el "Hessiano" (en problemas multiparametrales) analíticamente, los cuales se pueden incluir en la misma función `LogVSapp(...)` dada en línea 80 (i.e. `avScore[0]=diffLogVSapp(...)`). En el ejemplo subyacente no se han especificado ni el "Score" ni el "Hessiano", lo que indica al algoritmo `MaxBFGS(...)` que se deben usar las derivadas numéricas requeridas para calcular tales cantidades.

Con el uso de las funciones arriba mencionadas se definen las rutinas que resuelven las tareas uno y dos mencionadas en la planteación del problema. La rutina `SimX(const N,const lambda)` se encarga de simular N datos de la variable aleatoria cuya densidad es (1) y con el valor parametral λ . En esta función el truco usado en la línea 65 se hace únicamente para acelerar el proceso de simulación, ya que dada una sucesión ordenada de números uniformes $U_{(1)}, \dots, U_{(N)}$ la obtención del x que resuelve $F_X(x) - U_{(1)} = 0$ se utiliza como valor inicial para resolver $F_X(x) - U_{(2)} = 0$ y así sucesivamente. El re-ordenamiento original se regresa en la línea 72 gracias a que la función `sortindex` de `Ox` mantiene el ordenamiento original. La rutina encargada de la maximización esta dada en la línea 83 por `MV(const lambda0, const Pe, const x)`,

donde `lambda0` proporciona un valor inicial para el algoritmo (BFGS), `Pe` sirve para guardar el resultado y `x` provee con los datos muestrales.

Con fines de ilustración el programa principal (que comienza en la línea 23) simula los 1000 datos correspondientes al problema (línea 27) con valor parametral $\lambda = 1$. El histograma, así como la función original se dibujan mediante las funciones `DrawDensity` (que también puede dibujar un estimador suavizado de la densidad, aquí omitido) y `DrawXMatrix`, en esta última también se uso el hecho de que la función `fX(...)` admite forma matricial para evaluarla sobre un rango dado (`r=range(0,5,0.05)`). La función evaluada en el EMV también se dibuja en la línea 37 (nótese que se puede usar formato `LATEX` en las etiquetas). Figura 1 muestra el resultado salvado en formato `eps` mediante el uso de la función `SaveDrawWindow` en línea 39. La línea 40, sirve únicamente para editar la figura generada en `GiveWin`. Una explicación más detallada sobre los argumentos, de las funciones gráficas utilizadas se puede consultar en el archivo de ayuda antes mencionado.

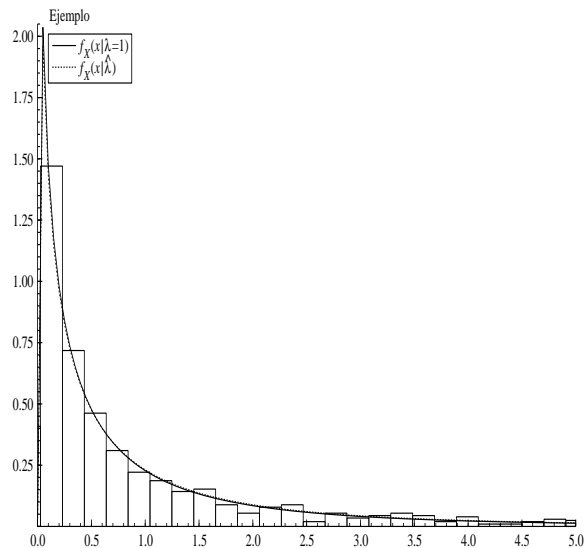


Figura 1: Ejemplo

La corrida del programa, i.e generada mediante el comando `oxl Ejemplo.ox` en una ventana de MS-DOS o en LINUX, genera lo siguiente

```
----- Ox at 17:34:20 on 05-May-2004 -----
Ox version 3.30 (Windows)
(C) J.A. Doornik, 1994-2003 This version
may be used for academic research and teaching only

Starting values parameters
  0.050000
gradients
  12781.
Initial function =
  -2015.41738792

Position after 7 BFGS iterations
Status: Strong convergence
parameters
  0.93654
gradients
  5.6843e-008
function value =
  -872.85419784

estimador para lambda=
  0.93654
Cannot show draw window!
```

La línea `Cannot show draw window!` se debe a que hemos invocado, en la línea 40, un comando que guarda un archivo `gwg`, el cual se puede editar únicamente en GiveWin, programa que se incluye en la versión no gratuita de Ox. Sin embargo, como se mencionó en el número anterior de DATOS, se puede incluir el paquete GnuDraw de Charles Bos que permite una interacción gráfica con GnuPlot que también es de distribución libre (sólo reemplazar la línea 8 por `#include <packages/gnudraw/gnudraw.h>`).

REFERENCIAS

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. AND FLANNERY, B. P. (1992). *Numerical recipes in C: The Art of Scientific Computing* ("2nd edition). Cambridge University Press.

PIESSENS, R. (1983). *Quadpack: A subroutine package for automatic integration*. Springer, Berlin.

DOORNIK, J.A. (2001). *Ox: An Object-Oriented Matrix Language (4th edition)*. Timberlake Consultants Press, London.

CRIBARI-NETO, F AND ZARKOS, S. G. (2003). *Econometric and Statistical Computing Using Ox. Computational Economics*. **21**, 277–295.

Ramsés Mena. Realizó sus estudios de licenciatura en Actuaría y de maestría en Estadística en la UNAM; acaba de recibir el grado de Doctor (Ph. D.) por la Universidad de Bath; sus especialidades son Estimación de procesos estocásticos, Series de tiempo y Estadística Bayesiana no paramétrica. Actualmente es investigador del IIMAS de la UNAM.

Galería de los grandes



Carl Friedrich Gauss
1777 - 1855

Gauss conformó el tratamiento de observaciones como una herramienta práctica



Navegando...

Humor

Sitios de interés en *internet*:

Más humor sobre Estadística:

Algunas ligas relacionadas con *software*, libros, manuales y demostraciones.

<http://www.buseco.monash.edu.au/~hyndman/jokes.htm>

<http://solon.cma.univie.ac.at/~neum/stat.html>
<http://members.aol.com/johnp71/javasta2.html>



Agradecemos el invaluable
apoyo de María Ochoa
(Unidad de Publicaciones
y Difusión, IIMAS)
en la edición de Datos.