

Automatización de la gestión de boletines de Econometría en la plataforma Moebius

Rey Míguez, Fernando (ferey@udc.es)

Rey Graña, Carla (carlarey@udc.es)

*Departamento de Economía Aplicada II
Universidad de A Coruña*

RESUMEN

En 2008 presentamos un original sistema de personalización de boletines para los alumnos de matemáticas de primer curso de Licenciatura de Administración y Dirección de Empresas.

Tras un largo proceso de desarrollo y de constante evolución, se ha diseñado un procedimiento que permite la generación automática de los boletines personalizados que deben realizar los alumnos y de sus soluciones, que se han integrado en la plataforma docente *Moebius* -<http://moebius.udc.es>- presentada en 2010.

Probada con éxito en la asignatura Matemáticas I de primer curso del Grado en Administración y Dirección de Empresas, se está implementando también en la asignatura Econometría de segundo curso de esta misma titulación, óptica desde la que abordamos esta novedosa automatización.

Palabras clave: Personalización; Evaluación; Plataforma docente

Área temática: Metodología y Didáctica.

ABSTRACT

In 2008 we presented an original system for personalizing the bulletins for first year mathematics students studying for the Business Management and Administration Degree.

After a long period of development and constant evolution, a procedure has been designed that automatically generates personalized bulletins to be filled out by the students together with their solutions. These form part of the teaching platform *Moebius* -<http://moebius.udc.es>- introduced in 2010. The system has been successfully trialled on the first year mathematics course of the Business Management and Administration Degree and is now being tested as part of the second year econometrics course for students studying for the same degree. It is from this perspective that we look at this novel form of automation.

Key words: Personalization; Evaluation; Teaching Platform

1. INTRODUCCIÓN

Hace más de tres años nos planteamos la posibilidad de preparar boletines personalizados para nuestros alumnos. Un primer paso en esta dirección derivó en la construcción de boletines individualizados para las asignaturas de matemáticas de primer curso de la Licenciatura de Administración y Dirección de Empresas, empleando una combinación por correspondencia entre la hoja de cálculo *Excel* y el procesador de textos *Word*, aplicaciones ambas del conocido paquete de software *Microsoft Office*. En [1] puede verse la construcción de un problema concreto empleando ésta técnica, así como dinámica básica de trabajo con los boletines propuestos.

Durante el curso 2009-10, coincidiendo con la implantación de los nuevos grados, nos dimos cuenta de la necesidad de construir una plataforma que facilitase la comunicación con el alumnado, sustituyendo al tradicional correo electrónico que estábamos empleando hasta el momento. Así, en el tercer año, nació nuestra plataforma *Moebius* construida ya íntegramente en *Software Libre*, que nos permitió agilizar la gestión informática e incorporar, sobre todo, algo de comodidad a una tarea que ya estaba alcanzando ciertas dimensiones. Pueden verse todos los detalles sobre la misma en [3].

Por la temática de nuestros problemas, que incluyen con frecuencia fórmulas matemáticas, el empleo de la combinación por correspondencia en *OpenOffice* en la construcción de ejercicios resultaba por momentos hartamente laboriosa y no estaba exenta de múltiples restricciones; necesitábamos emplear un procesador de texto matemático que nos diese mayor libertad y comodidad, y que permitiese un procesamiento dinámico de fórmulas. Con el uso del procesador matemático $\text{\LaTeX}$ y una combinación por correspondencia de construcción propia empleando *AWK* y *Bash* (véase epígrafe 2.2 en [3]), empezamos a obtener resultados alentadores.

Se crearon múltiples algoritmos para facilitar la escritura de fórmulas matemáticas en $\text{\LaTeX}$, plantillas para enunciados y soluciones, filtros estándar, funciones básicas, etc. También un amplio abanico de procedimientos matemáticos que se integran convenientemente en el conjunto: una robusta construcción modular que permite generar de forma automática y muy rápida los boletines para nuestros alumnos y las soluciones para los profesores.

Seguidamente y haciendo especial hincapié en las diferencias comparativas con el trabajo [1] abordamos su revisión, particularizando en los problemas de los boletines desde la perspectiva de la asignatura Econometría de 2º curso del Grado de ADE.

2. UN PROBLEMA TIPO

La estimación de un modelo econométrico básico se efectúa utilizando el método de mínimos cuadrados ordinarios (MCO) porque permite obtener unos estimadores que satisfacen todas las propiedades deseables.

Genéricamente la ecuación que representa un modelo básico es:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

Se asume que ε_t es una variable aleatoria normal, de esperanza matemática y covarianzas nulas, con varianza constante σ^2 .

Dado que nuestro propósito es generar problemas similares partiendo de uno concreto y a pesar de que en un modelo clásico se asume que los valores de los regresores son no estocásticos, para que cada alumno dispusiese de una tabla de observaciones diferentes se han efectuado variaciones proporcionales de sus observaciones muestrales.

También se utilizó este procedimiento con los valores del regresando que es una variable aleatoria que sigue una distribución $N(0, \sigma)$, y como consecuencia las observaciones proporcionales seguirán una distribución $N(0, \alpha\sigma)$, siendo α el factor

de proporcionalidad. Por esta razón, para generar los valores de la variable endógena se ha utilizado la expresión:

$$\bar{y}_t = \alpha y_t + v_t, \quad (2)$$

donde v es una variable aleatoria $N(0, \alpha S)$, siendo S el valor de la desviación típica estimada de la perturbación obtenida al estimar por MCO el modelo original.

Para simular los valores de v , tomando como base el teorema central del límite, se ha utilizado la expresión:

$$v_t = S^* \sqrt{\frac{12}{K}} \left(\sum_{i=1}^K r_i - \frac{K}{12} \right) \quad (3)$$

donde $S^* = \alpha S$ y las variables aleatorias $r_i, i = 1, \dots, K$ son independientes e idénticamente distribuidas definidas en el intervalo $(0, 1)$. Para K un valor natural razonable sería 100, sin embargo hemos elegido $K = 108$ ya que de esa forma la expresión (3) se reduce a:

$$v_t = \frac{S^*}{3} \left(\sum_{i=1}^{108} r_i - 9 \right).$$

A continuación deben establecerse unos filtros, ya que no todos los modelos generados serán válidos. Pudiera ser -por ejemplo- que algún valor $\bar{y}_t$ en (2) fuese negativo, que los estimadores en la ecuación (1) tengan signos incorrectos desde el punto de vista teórico, que las varianzas estimadas de los estimadores sean excesivamente pequeñas, que la matriz $X'X$ genere problemas computacionales, etc; todos estos casos habría que desecharlos.

El establecimiento de todos estos filtros -que se hace vía programación- es relativamente sencillo y sería inviable su implementación en una hoja de cálculo salvo que quisiéramos obtener un número reducido de ellos, aún así el proceso sería manual y muy laborioso.

Los filtros pueden ir más allá y seleccionarse modelos que tengan, por ejemplo, un coeficiente de desigualdad de Theil en unos márgenes determinados, un cam-

bio estructural en un determinado período, una selección de regresores con idénticos resultados independientemente del método elegido, etc. Las posibilidades son amplísimas y como veremos más adelante, el tiempo necesario en implementarlos muy reducido.

3. LA PLANTILLA DE SOLUCIONES

Los cajetines para que los alumnos escriban las soluciones varían según el problema, pero es frecuente que se haga algún contraste, por lo que se han preparado módulos de impresión específicos para agilizar esta tarea. Por ejemplo, para el contraste de relevancia individual y efecto positivo de la variable población en un modelo que explica el comportamiento de las ventas de automóviles, se tiene el siguiente cajetín de soluciones:

hipótesis nula	hipótesis alternativa	estadístico de prueba
$H_0 : \beta_2 = 0$	$H_1 : \beta_2 > 0$	$t_2 = \frac{b_2}{S_{b_2}}$

valor estadístico t	punto crítico	conclusión
22.7622	1.895	rechazo H_0

que el alumno debe escribir en una caja del tipo:

hipótesis nula	hipótesis alternativa	estadístico de prueba

valor estadístico t	punto crítico	conclusión

Otros cajetines estándar diseñados permiten mostrar el modelo estimado:

modelo estimado	$\hat{y}_t = 2,3651 + 0,1234x_{1t} + 0,0560x_{2t}$ $(0,4541) \quad (0,0018) \quad (0,0025)$
-----------------	--

En ocasiones, para facilitar la tarea, se le proporcionan alguno de los valores intermedios como guía.

4. LOS AUTOMATISMOS

4.1. Generación de boletines

Creada una asignatura dentro de la plataforma (véase epígrafe 4 en [3]) y tras seleccionar los problemas deseados dentro de la colección existente, la invocación de comandos diseñados específicamente, realiza las tareas necesarias con un esquema similar al presentado en [3]. El comando básico de colocación de boletines en las cuentas de los alumnos es:

bopack *asig* *nº*

donde *asig* es la asignatura seleccionada y *nº* es el número de boletín en cuestión.

Este potente comando genera todos los valores numéricos necesarios de cada uno de los problemas del boletín. En base a ellos, crea el código $\text{\LaTeX}$ del enunciado y la plantilla de soluciones. Otros módulos crean varios tipos de cabeceras (estándar, gráficas, de saludo, numéricas, etc), marcas de agua, pies de página,... con los datos correspondientes a cada alumno y su profesor. Todos los archivos generados van identificados y se combinan apropiadamente para obtener cada uno de los boletines fuente $\text{\LaTeX}$ que se compilan individualmente, obteniéndose así un boletín para cada alumno que se envía a su cuenta y del que ya puede disponer para su realización.

4.2. Generación de soluciones para su corrección

La corrección de todos estos boletines sería inviable si no facilitamos las soluciones a los profesores; cada docente recibe en su cuenta los boletines resueltos de sus alumnos tras la invocación del comando:

boenunsol *asig n°*,

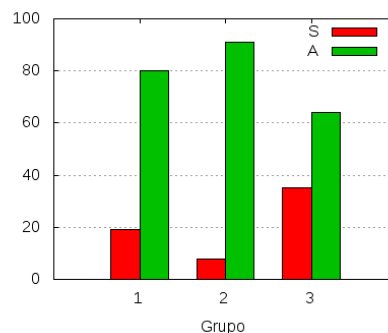
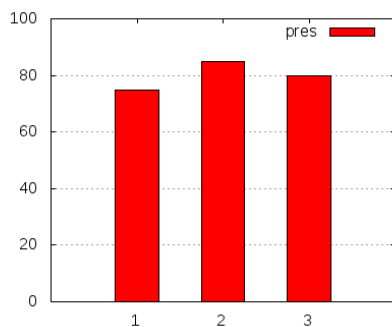
además de todos los enunciados por si fuese necesario realizar alguna comprobación. Corregidos los boletines, cada profesor cubrirá la plantilla de notas de su/s grupos en una sencilla hoja *Excel* que también se encuentra en su cuenta y que enviará al *webmaster*.

4.3. Envío de notas y soluciones a los alumnos

La orden:

bosolalumnos *asig n°*

es la encargada de colocar las notas y soluciones correctas de cada uno de los boletines en las cuentas de los alumnos que lo hayan realizado. También actualiza la base de datos con las notas de cada grupo en las cuentas de los profesores y les facilita una completa y útil estadística de los resultados obtenidos. Como muestra, dos gráficas comparativas por grupos del primer boletín de este curso.



En la primera se representa el porcentaje de alumnos que lo han realizado y en la segunda el porcentaje de aprobados y suspensos. Es fácil apreciar el mayor interés y mejores resultados obtenidos por el grupo 2.

4.4. Generación de avisos: el tablón

De forma automática y asociada a las órdenes anteriores, un tablón de anuncios informa al alumno de la disponibilidad tanto de los enunciados de los boletines como de sus notas y las soluciones correctas.

```
/*      \\\|/
      (o o)
+-----o00-{ }-00o-----+
|
|  BOLETINES
|      -> Disponible boletín a entregar 2.
|      -> Disponibles notas y soluciones del boletín 1.
|
|  CONTROLES
|      -> Disponibles notas y soluciones del control 1.
|
|  TEMAS
|      -> Disponibles : 1-6
|
|                                     ·Moebius·
+-----+*/
```

El aviso anterior se corresponde con la situación en el momento de redactar el presente trabajo. Adicionalmente, el tablón proporciona otro tipo de información como los capítulos disponibles para descarga o sobre los controles periódicos que se realizan en el aula como parte de la evaluación.

5. LOS RECURSOS

Todos los automatismos precedentes carecerían de interés si su realización fuera costosa. Al contrario, el diseño modular desarrollado permite la ejecución de los

procesos con gran precisión y rapidez.

Consideremos un problema¹ cuyo objetivo es que el alumno compruebe, entre otras cosas, que la omisión de una variable relevante intensamente relacionada con las incluidas puede ocasionar que tanto la varianza estimada de la perturbación como las varianzas estimadas de los estimadores disminuyan:

“Se dispone de la siguiente muestra de valores para las variables consumo y , renta x_1 , y riqueza x_2 para 10 familias:

familia	y	x_1	x_2
1	3.00	6.00	18.00
2	4.00	11.00	32.00
3	5.00	10.00	30.00
4	6.00	18.00	55.00
5	7.00	14.00	44.00
6	9.00	18.00	53.00
7	8.00	24.00	69.00
8	10.00	26.00	75.00
9	12.00	34.00	111.00
10	14.00	46.00	133.00

Se pide:

1. Calcular el determinante de la matriz de correlaciones.
2. Contrastar la relevancia individual y conjunta de las variables explicativas del modelo.
3. Eliminar de el modelo la variable x_2 , calcular la variación en la SCE , en la varianza estimada de la perturbación y en la varianza estimada de los estimadores.”

Supongamos que necesitamos -como es nuestro caso- generar exactamente 189 boletines con este problema para nuestros alumnos, todos ellos con similares características a partir del anterior, por lo que hemos impuesto los siguientes filtros:

¹véase completo en el anexo.

- Los estimadores MCO son positivos y mayores que una milésima.
- Las desviaciones típicas estimadas de los estimadores Sb_0 , Sb_1 , Sb_2 son mayores que una milésima.
- Exclusión de matrices $X'X$ que puedan ocasionar problemas numéricos.
- $r_{12} > 0,99$, que garantiza un elevado grado de multicolinealidad, apartado (a).
- En el apartado (c), al eliminar la variable x_2 :
 - Se incrementa SCE .
 - La varianza estimada de la perturbación disminuya.
 - La varianza estimada del estimador disminuya.
 - x^{11} disminuya.

La tabla siguiente muestra los recursos empleados tras esta rigurosa selección:

archivos de trabajo	2498
tiempo de cálculo*	0.9
números aleatorios generados	649133
regresiones	1201
archivos compilados	189
líneas látex compiladas	84105
tiempo de compilación*	17.57
tiempo total*	18.47

*tiempo medido en segundos

El ordenador utilizado incorpora un potente procesador *Intel Core i7-920* a $2.67GHz$. Sin embargo, no se necesita un ordenador tan veloz; repetido el proceso anterior con un portátil estándar con un procesador *Intel Core i3* a $2.2GHz$ se ha obtenido un incremento

de aproximadamente un 50 % en el tiempo total empleado, pero en los márgenes de tiempo que nos movemos esto es irrelevante.

Queremos llamar la atención de la magnitud de la tarea; se han generado casi seiscientos cincuenta mil números aleatorios, realizado más de 1200 regresiones y se han compilado aproximadamente 84000 líneas de código $\text{\LaTeX}$ -equivalente a unos 309 metros de longitud- todo ello en unos 18 segundos. Si el boletín lo completamos con un par de problemas más, hablaríamos en cifras redondas de un minuto en total; la velocidad es extraordinaria.

Cabe mencionar también que la calidad tipográfica es comparable a la de una editorial académica gracias al empleo de $\text{\LaTeX}$, formato en el que está elaborado el presente documento.

Faltaría por incluir en la tabla anterior el tiempo de transferencia y colocación en las cuentas de los alumnos de los boletines, pero esta tarea es prácticamente instantánea.

6. LOS NÚMEROS ALEATORIOS

Uno de los elementos clave para el correcto funcionamiento de nuestra plataforma es el generador de números aleatorios. Inicialmente se empleó el facilitado por el lenguaje *AWK*; *srand()* para la semilla y *rand()* para la generación sucesiva. Para tener un mayor control sobre los mismos, se usó posteriormente el algoritmo AS 183 propuesto por Wichmann & Hill [5]. En la actualidad ha sido sustituido por el propuesto por los mismos autores en [6], un GCL (Generador Congruencial-Lineal) combinado de cuatro ciclos en vez de los 3 del anterior. Las cuatro semillas necesarias para comenzar la secuencia aleatoria con un período aproximado de 2^{121} se establecen mediante un algoritmo propio dependiendo de la asignatura, año y número de alumno, por lo que un cambio en cualquiera de los anteriores parámetros, conduce a una secuencia aleatoria diferente que genera un boletín con unos valores numéricos impredecibles pero perfectamente controlados.

7. GESTIONANDO VARIAS ASIGNATURAS

La gestión automática de los boletines se ha integrado a la perfección dentro de la plataforma *Moebius*, lo se traduce en la posibilidad de administrar simultáneamente varias asignaturas dado que cada una de ellas tiene su propia identificación. Obviamente antes es necesario crear las cuentas de alumnos y profesores, fijar el número de boletines, de grupos, etc. (véase epígrafe 4 en [3]), en el que el proceso de creación se ha mejorado notablemente, traducándose para al *webmaster* en facilidad y velocidad de gestión.

Por tanto, la ejecución de sencillos comandos por parte del *webmaster* permitiría tener a disposición de los alumnos boletines individualizados de características similares a los que los accederían a través de la red y, por su parte, los profesores dispondrían de las soluciones para posibilitar su corrección. A todo esto ha de añadirse la eficaz gestión de las soluciones y los resultados, todo ello de forma rápida, precisa, automática y garantizando la exclusividad de cada uno de los problemas de los boletines.

8. UNA NOTA FINAL

En la actualidad estamos trabajando en la automatización de los tests personalizados que se abordaron en [2]. El trabajo está bastante avanzado y los resultados obtenidos son prometedores.

9. REFERENCIAS BIBLIOGRÁFICAS

- [1] REY MÍGUEZ, F.; PEDREIRA ANDRADE, L. P. (2008). "Personalización de problemas: hacia un nuevo entorno de evaluación", en *Recta@2008 Vol: Actas_16 Issue 1*: 612.
- [2] REY MÍGUEZ, F.; PEDREIRA ANDRADE, L. P. (2009). "Tests aleatorios: hacia un nuevo entorno de evaluación", en *Rect@2009 Vol: Actas_17 Issue 1*: 610.

- [3] REY MÍGUEZ, F.; PEDREIRA ANDRADE, L. P. (2010). “Moebius: una plataforma docente basada en Software Libre”, en Rect@2010 Vol: Actas_18 Issue 1: 110.
- [4] SÁNCHEZ PRIETO, S.; GARCIA POBLACIÓN, O. (2008). LINUX, Guía práctica. Ed. Ra-Ma.
- [5] WICHMANN, B.; HILL, I. (1982). “Algorithm AS 183: An efficient and portable pseudo-random number generator”. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol.31, N°2. Blackwell.
- [6] WICHMANN, B.; HILL, I. (2006). “Generating good pseudo-random numbers”. Computational Statistics & Data Analysis. Volume 51 Issue 3, December, 2006 Journal of de Royal Statistical Society. Series C (Applied Statistics), Vol.51, N°3. Elsevier.
- [7] OETIKER, T y otros. (2009). “The Not So Short Introduction to L^AT_EX 2_ε”. Descarga en <http://ctan.org/tex-archive/info/lshort/english/lshort.pdf>
- [8] VIDAL CORTÉS, J. (2002). “The GNU Awk User’s guide”. En <http://www.gnu.org/software/gawk/manual/gawk.html>

ANEXO

Presentamos en este anexo el problema completo abordado en el presente trabajo. Comenzamos por el enunciado, seguido de las soluciones y finalmente el cajetín de soluciones vacío en el que el alumno debe escribir las soluciones. Se ha reducido el espacio en blanco de los cajetines para ocupar menos espacio. En la práctica los huecos son mucho más amplios par contener cómodamente los números y fórmulas.

A.1. Enunciado

Se dispone de la siguiente muestra de valores para las variables consumo y , renta x_1 , y riqueza x_2 para 10 familias:

familia	y	x_1	x_2
1	3.00	6.00	18.00
2	4.00	11.00	32.00
3	5.00	10.00	30.00
4	6.00	18.00	55.00
5	7.00	14.00	44.00
6	9.00	18.00	53.00
7	8.00	24.00	69.00
8	10.00	26.00	75.00
9	12.00	34.00	111.00
10	14.00	46.00	133.00

Se pide:

1. Calcular el determinante de la matriz de correlaciones.
2. Contrastar la relevancia individual y conjunta de las variables explicativas del modelo.
3. Eliminar de el modelo la variable x_2 , calcular la variación en la SCE, en la varianza estimada de la perturbación y en la varianza estimada de los estimadores.

A.2. Soluciones

1.	modelo estimado $\hat{y}_t = 2,0628 + 0,0741x_{1t} + 0,0678x_{2t}$ $(0,7433) \quad (0,3003) \quad (0,1005)$
----	---

$S_{x_1x_2}$	S_{x_1}	S_{x_2}	r_{12}	$ R_x $
398.4000	11.5763	34.6035	0.9946	0.0108

Conclusión	Elevado grado de multicolinealidad
------------	------------------------------------

2. CONTRASTES DE RELEVANCIA

a) Renta

hipótesis nula	hipótesis alternativa	estadístico de prueba
$H_0 : \beta_1 = 0$	$H_1 : \beta_1 \neq 0$	$t_1 = \frac{b_1}{S_{b_1}}$

valor estadístico t	punto crítico	conclusión
0.2469	2.3646	no rechazo H_0

b) Riqueza

hipótesis nula	hipótesis alternativa	estadístico de prueba
$H_0 : \beta_2 = 0$	$H_1 : \beta_2 \neq 0$	$t_2 = \frac{b_2}{S_{b_2}}$

valor estadístico t	punto crítico	conclusión
0.6747	2.3646	no rechazo H_0

c) Conjunta

hipótesis nula	hipótesis alternativa	estadístico de prueba
$H_0 : \underline{\beta} = 0$	$H_1 : \underline{\beta} \neq 0$	$F = \frac{T - k - 1}{k} \frac{R^2}{1 - R^2}$

valor de F	punto crítico	conclusión
39.0602	4.7374	rechazo H_0

3. En este apartado el *modelo1* se refiere al modelo inicial (completo) y el *modelo2* al modelo resultante de eliminar la variable x_2

	modelo 1	modelo 2	diferencia	CONCLUSIÓN
SCE	9.1776	9.7745	0.5969	SCE $\nearrow$
S^2	1.3111	1.2218	-0.0893	la varianza est. de la perturbación $\searrow$
$S^2_{b_1}$	0.0902	0.0009	-0.0893	la varianza est. del estimador $\searrow$
x^{11}	0.0688	0.0007	-0.0680	x^{11} $\searrow$

A.3. Cajetines de soluciones

1.	<table border="1"> <tr> <td>modelo estimado</td><td>$\hat{y}_t =$</td></tr> </table>	modelo estimado	$\hat{y}_t =$
modelo estimado	$\hat{y}_t =$		

$S_{x_1x_2}$	S_{x_1}	S_{x_2}	r_{12}	$ R_x $

Conclusión	
------------	--

2. CONTRASTES DE RELEVANCIA

a) Renta

hipótesis nula	hipótesis alternativa	estadístico de prueba

valor estadístico t	punto crítico	conclusión

b) Riqueza

hipótesis nula	hipótesis alternativa	estadístico de prueba

valor estadístico t	punto crítico	conclusión

c) Conjunta

hipótesis nula	hipótesis alternativa	estadístico de prueba

valor de F	punto crítico	conclusión

- d) En este apartado el *modelo1* se refiere al modelo inicial (completo) y el *modelo2* al modelo resultante de eliminar la variable x_2

	modelo 1	modelo 2	diferencia	CONCLUSIÓN
SCE				
S^2				
$S_{b_1}^2$				
x^{11}				