

La simulación y la enseñanza de la estadística. Casos prácticos

por
PEDRO GRIMA y ALEXANDRE RIBA
Departamento de Estadística e Investigación Operativa
Universidad Politécnica de Catalunya

RESUMEN

En este trabajo se exponen algunas ideas sobre cómo aprovechar las posibilidades de simulación que presentan los paquetes de software estadístico en el contexto de la enseñanza de la estadística. Se presentan casos concretos en temas tales como la ilustración de conceptos estadísticos básicos, la resolución de problemas de cálculo de probabilidades y aspectos relacionados con la variabilidad en los procesos de fabricación.

Los casos presentados se resuelven, a título de ejemplo, con ayuda del paquete MINITAB.

Palabras clave: Simulación, distribución de estadísticos muestrales, cálculo de probabilidades, paquete de software estadístico

Clasificación AMS: 65C05, 68020, 6201.

1. INTRODUCCIÓN

La difusión de los ordenadores y de los paquetes de software estadístico, además de aumentar la posibilidad de tratar grandes volúmenes de datos y de facilitar la aplicación de métodos estadísticos que exigen cálculos muy laboriosos, abre también un gran abanico de posibilidades al uso de las técnicas de simulación como ayuda a la docencia y a la resolución de problemas.

El tema no es nuevo. Ya en los números 122 y 123 de esta revista publicados a finales del 89 y principios del 90, y dedicados monográficamente a la enseñanza de la estadística en España se comenta por parte de distintos profesores la importancia y las posibilidades de la simulación en la enseñanza de esta disciplina.

En este sentido, Novales (1989) describe la importancia y las posibilidades de la simulación en la Estadística aplicada a la Economía y la Econometría. Santisteban (1989) comenta el uso de la simulación en la Estadística aplicada a la Psicología y, en el ámbito de las Escuelas Técnicas, en el artículo de Peña, Prat y Romero (1990) se comenta como en la E.T.S. de Ingenieros Industriales de Barcelona ya se empezaba entonces a utilizar el programa MINITAB para verificar empíricamente, mediante simulación, los resultados más interesantes de la estadística matemática.

Más recientemente, Peña (1992) describe las posibilidades de la simulación en distintos ámbitos (resolución aproximada de problemas de probabilidad, simulación de distribuciones, ...) comentando sus posibilidades pedagógicas.

Este artículo pretende ilustrar, a través de algunas reflexiones basadas en la experiencia propia y de un conjunto de ejemplos seleccionados por el interés que despiertan en los estudiantes, como la aplicación de técnicas de simulación puede resultar útil para mejorar los métodos de enseñanza de la estadística. Asimismo, los ejemplos que se incluyen también pueden servir de posible fuente de inspiración para incorporar ejercicios de simulación en las clases prácticas, especialmente en carreras de tipo técnico.

Dichos ejemplos, muchos de ellos extraídos de las clases prácticas de estadística que se realizan en la E.T.S. de Ingenieros Industriales de Barcelona y en la Facultad de Matemáticas y Estadística de la Universidad Politécnica de Cataluña (Ginebra et al., 1995), se presentan divididos en 3 apartados. En el primero de ellos se incluyen 2 ejemplos en los que se pone de manifiesto como la simulación puede ser útil tanto para poner de manifiesto comportamientos estadísticos importantes (distribución de la media y de la varianza muestral) como para ilustrar qué es un contraste de hipótesis a través de la elaboración, por parte de los propios estudiantes, de la distribución de referencia para la aplicación del test de Kolmogorov.

El segundo apartado contiene algunos curiosos problemas de cálculo de probabilidades que presentan una cierta dificultad si se pretenden resolver por los métodos clásicos, y que aquí se presentan resueltos por simulación, comentando las ventajas que este procedimiento puede aportar a la docencia en muchos casos.

Finalmente se incluyen dos ejemplos de aplicación de la simulación para ilustrar fenómenos relacionados con la aplicación de la estadística al control y la mejora de la calidad.

En todos los casos se utiliza el paquete de software estadístico MINITAB, que consideramos muy adecuado para la docencia. Esencialmente consiste en una hoja de cálculo en la que se colocan los datos, y en un conjunto de comandos y subcomandos que permiten ir realizando los análisis que se deseen. Pueden también encadenarse una serie de comandos formando lo que se denomina una macro, lo cual permite ejecutarlos de forma conjunta automáticamente tantas veces como se desee.

Con el fin de que sirva como sugerencia para los lectores interesados, y también para poner de manifiesto su sencillez, se incluyen algunas de las macros con las que se han realizado las simulaciones que se comentan en este artículo.

2. ILUSTRACIÓN DE CONCEPTOS ESTADÍSTICOS BÁSICOS

La experiencia nos demuestra que algunos conceptos estadísticos resultan difíciles de asimilar por parte de los estudiantes. Las razones de esta dificultad de aprendizaje pueden ser varias. Watts (1991) señala que la mayor dificultad a la que se enfrentan los estudiantes es que los conceptos estadísticos son básicamente abstractos. Por esta razón, quizá no se han interiorizado bien ideas clave como la de variable aleatoria o error experimental, o quizá la ley resulta chocante y poco intuitiva, o por otros motivos (la falta de "inspiración" del profesor puede ser, sin duda, uno de ellos).

En cualquier caso, el poder constatar, el **ver**, que la ley enunciada efectivamente se cumple, y que es lo que hay detrás del concepto expuesto, sin duda ayudara a comprenderlo y a asimilarlo.

2.1. Distribución de la media y la varianza muestral

El hecho de que la media muestral presente una distribución normal bajo determinadas condiciones de distribución de la población (1) y de tamaño de muestra, suele ser considerado por los estudiantes como un resultado sorprendente. Este resultado se puede ilustrar en clase de teoría (queremos decir en una clase convencional, sin ordenadores pero con ayuda de un proyector de transparencias) mostrando los gráficos de la distribución de la puntuación media obtenida al lanzar un determinado número de dados, según sea dicho número. Este ejemplo permite observar como la distribución de las medias se va aproximando a una Normal a medida que aumenta el número de dados lanzados. En Prat, Tort-Martorell, Grima y Pozueta (1994) se incluyen gráficos obtenidos por simulación que ponen de manifiesto este fenómeno.

Y una buena forma de asegurar la asimilación de esta importante idea es que los propios estudiantes comprueben por simulación, en las clases prácticas, que cuanto se ha dicho en clase de teoría efectivamente se cumple.

La simulación puede realizarse generando muestras de números aleatorios (2) con determinada distribución de probabilidad, calculando su media y repitiendo el proceso de forma iterativa almacenando las medias obtenidas. Una vez terminada la simulación, puede realizarse el histograma de las medias y comprobar la forma de su distribución.

La macro de la figura 1 genera muestras de 15 elementos de una distribución uniforme discreta definida entre 1 y 6 (15 lanzamientos de un dado, por ejemplo) y va acumulando sus medias. Ejecutada la macro 1000 veces puede representarse la distribución de las medias muestrales con el resultado que se indica en la figura 2.

(1) Condiciones que en ciertos contextos y campos de aplicación de la Estadística pueden ser obviadas. Citando a Box, Hunter y Hunter (1978): "Ciertamente en general para las distribuciones usuales. No es verdad para ciertos juguetes matemáticos (v.g.: distribución de Cauchy) de la que no debemos preocuparnos" (pág. 66).

(2) En rigor deberíamos decir "números pseudo-aleatorios", ya que las realizaciones de variables "aleatorias" están determinadas algorítmicamente y, por lo tanto, no son estrictamente aleatorias, aunque superen los test de aleatoriedad. En Devroye (1986) o Lewis (1975) puede encontrarse la formulación teórica de generadores de números pseudo-aleatorios. En este trabajo se utiliza la denominación de "aleatorio" por ser la que habitualmente se usa en estos contextos.

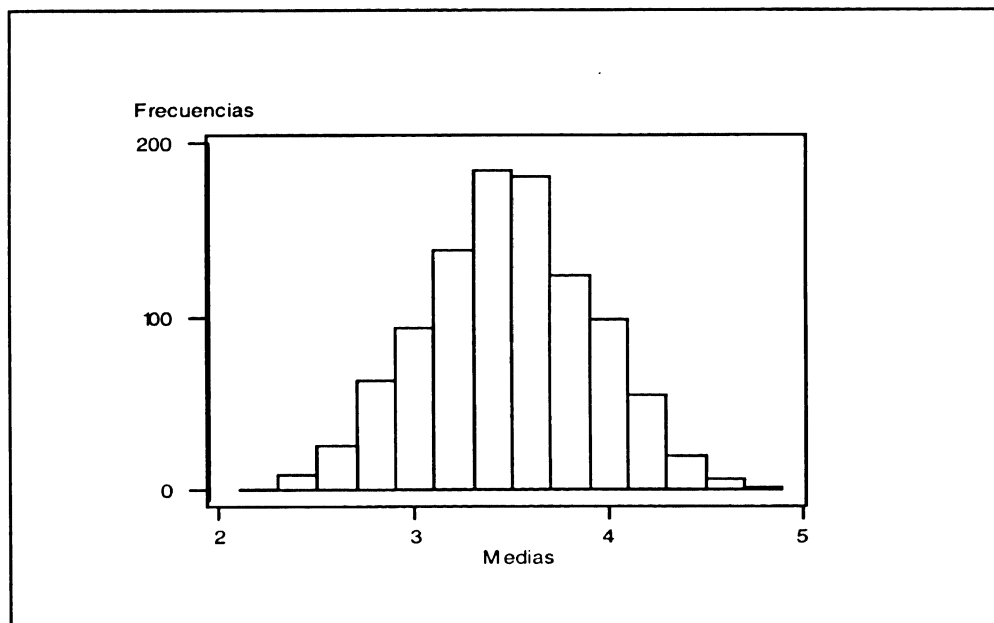
Figura 1

MACRO PARA SIMULAR LA OBTENCIÓN DE MUESTRAS DE UNA DISTRIBUCIÓN EQUIPROBABLE DEFINIDA ENTRE 1 Y 6, CALCULANDO Y ALMACENANDO LA MEDIA DE CADA MUESTRA

```
# Previamente hay que definir: k2=1  
RANDOM 15 C1;      # Genera 15 números aleatorios ..  
INTEGER 1 6.      # ... de una dist. equiprobable entre 1 y 6  
MEAN C1 K1        # Coloca la media de los valores de C1 en K1  
LET C2(K2)=K1     # Va acumulando las medias en C2  
LET K2=K2+1  
END
```

Figura 2

HISTOGRAMA CORRESPONDIENTE A LAS 1000 MEDIAS OBTENIDAS EN LA MACRO DE LA FIGURA 1



La prueba puede realizarse también para otros tamaños de muestra y/o otras distribuciones de probabilidad (puede incluso dejarse al criterio del estudiante el elegir otras distribuciones de la población u otros tamaños de muestra). En todos los casos se puede comparar, además de otros aspectos interesantes, la relación entre las varianzas de la media muestral y la varianza poblacional.

También, por un procedimiento similar, puede estudiarse la distribución de la varianza muestral y observar que en este caso no aparece la distribución Normal. Mediante la macro de la figura 3 se generan muestras aleatorias de tamaño 9 de una población $N(0;1)$ y se calcula su varianza. Realizando 1000 iteraciones se obtienen otras tantas varianzas muestrales de las cuales puede realizarse su histograma (figura 4).

Figura 3

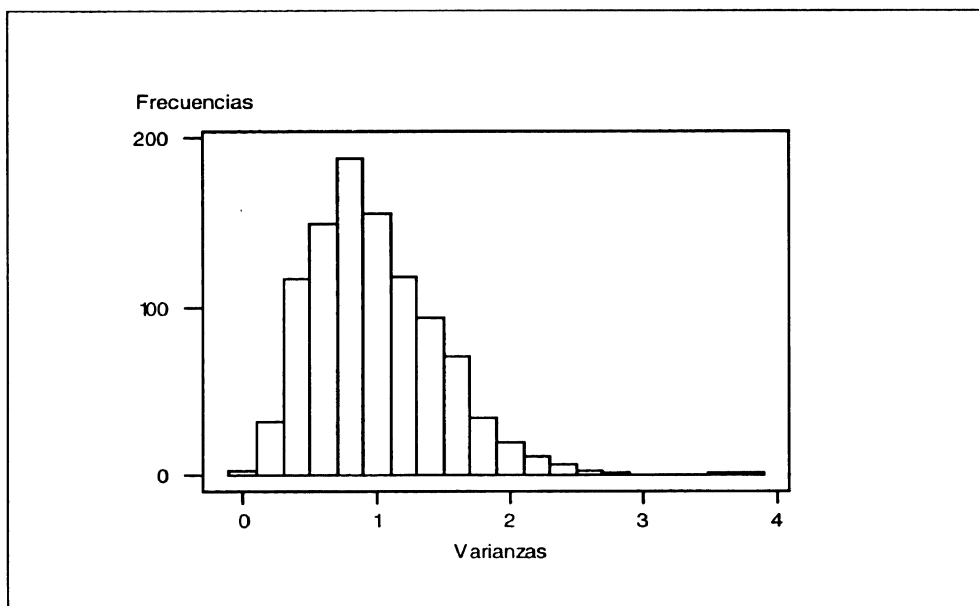
MACRO PARA SIMULAR LA OBTENCIÓN DE MUESTRAS DE UNA DISTRIBUCIÓN NORMAL ESTANDARIZADA, CALCULANDO Y ALMACENANDO LA MEDIA Y LA VARIANZA DE CADA MUESTRA

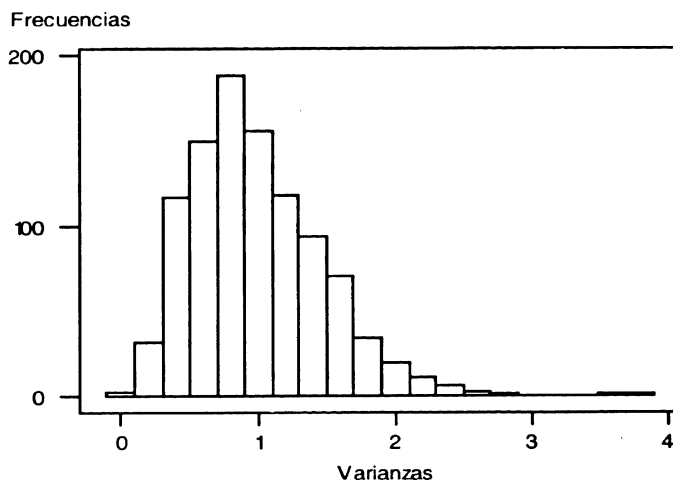
```
# Previamente hay que definir: k3=1

RANDOM 9 C1          # Genera 9 valores de una N(0;1)
MEAN C1 K1           # Asigna a K1 el valor de la media
STDEV C1 K2          # Asigna a K2 el valor de la desv. tipo
LET C2(K3)=K1        # Acumula las medias en C2
LET C3(K3)=K2*K2      # Acumula las varianzas en C3
LET K3=K3+1          # Suma 1 al contador
END
```

Figura 4

HISTOGRAMA DE LAS VARIANZAS OBTENIDAS MEDIANTE LA MACRO DE LA FIGURA 3





Además, recordando que:

$$s^2 \approx \frac{\sigma^2}{(n-1)} \chi^2_{n-1}$$

puede deducirse que multiplicando la columna de varianzas por 8 (es decir, $n-1$) se obtienen valores de una distribución chi-cuadrado con 8 grados de libertad ya que en este caso $\sigma^2=1$.

Resulta interesante que los estudiantes comprueben que, “tal como indica la teoría”, la media de los valores así obtenidos está en torno a 8 y su desviación tipo en torno a 4. (Recuérdese que $E(\chi^2_{n-1}) = n-1$, y $V(\chi^2_{n-1}) = 2(n-1)$). También puede resultar un buen ejercicio el tabular estos datos y comparar sus frecuencias de aparición con los valores que aparecen en las tablas χ^2 con 8 grados de libertad.

2.2. Ejemplo de contraste de hipótesis: El test de Kolmogorov

Un concepto sin duda importante, y que no siempre se asimila con facilidad, es el de contraste de hipótesis. Una forma de ayudar a los estudiantes a “visualizar” este método de razonamiento y, por tanto de facilitar su comprensión, es proponerles que ellos mismos deduzcan por simulación la distribución de referencia (o algunos de sus valores) para alguno de los tests estadísticos que formen parte del curso.

La metodología es muy sencilla, consiste en ir generando muestras de un cierto tamaño y calculando D_{\max} para cada una de ellas. Repetida la operación un determinado número de veces, puede verse cual es la distribución de este estadístico y cuales son los valores que, por ejemplo, son sobrepasados el 5 % y el 1 % de las veces, que son los que normalmente aparecen en las tablas.

Hecho esto, no resulta difícil entender lo que se hace al aplicar el Test de Kolmogorov: comparar el valor de D_{\max} que corresponde a la muestra analizada con la distribución que presentaría ese valor si la muestra proviniera realmente de la distribución que se le supone. Esta claro que según sea de "raro" el estadístico de prueba en su distribución de referencia, la decisión a tomar respecto a la hipótesis nula será una u otra

En la figura 5 se reproduce la macro que se ha construido para encontrar los valores correspondientes a $n=10$ en el caso de que la distribución de H_0 sea normal y se conozcan sus parámetros. Realizando 1000 iteraciones se han obtenido los valores 0.4154 para $\alpha=0.05$ y 0.4872 para $\alpha=0.01$.

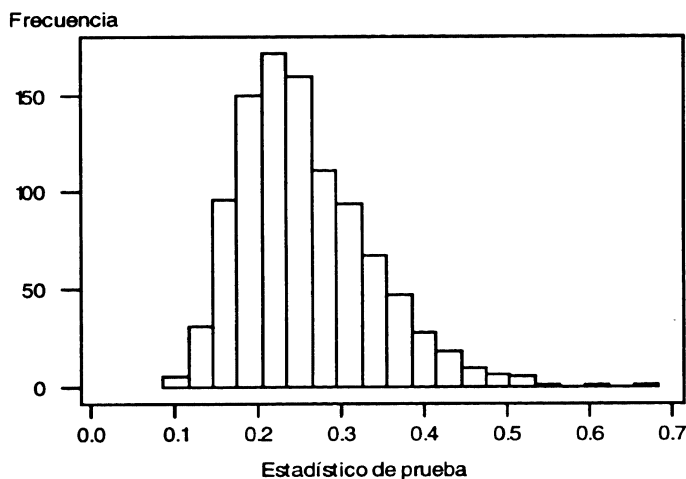
Figura 5

MACRO PARA DETERMINAR LOS VALORES CORRESPONDIENTES A $N=10$ DEL TEST DE KOLMOGOROV, CON PARÁMETROS CONOCIDOS INDEPENDIENTES DE LA MUESTRA (KOLMOGOROV-MASSEY)

```
# Previamente es necesario introducir C10 = 0.1, 0.2, ... 1
# y C11 = 0, 0.1, 0.2, ... 0.9, que son los valores de  $F_*(x_i)$  y
#  $F_*(x_{i-1})$  cuando  $n=10$ .
# También es necesario definir K4=0 (Puesta a 0 del contador)
#
RANDOM 10 C1; # Genera 10 núm. aleatorios ...
NORMAL 10 2. # ... de una distribución N (10;2)
SORT C1 C2      # Ordena los valores generados
CDF C2 C3;      # Calcula los valores de  $F_t(x_i)$  ...
NORMAL 10 2. # ... de acuerdo con la distribución de  $H_0$ 
LET C4=C3-C10 # Calcula las diferencias entre  $F_t(x_i)$  y  $F_*(x_i)$ 
LET C5=C3-C11 # Calcula las diferencias entre  $F_t(x_i)$  y  $F_*(x_{i-1})$ 
STACK C4 C5 C6 # Une todas las diferencias en una sola columna
ABSOL C6 C7      # Toma el valor absoluto de las diferencias
MAX C7 K3        # Determina el valor máximo de las diferencias
LET K4=K4+1      # Añade 1 al contador
LET C8(K4)=K3    # Acumula la diferencia máxima
PRINT K4
END
```


Figura 6

DISTRIBUCIÓN DE ESTADÍSTICO DE PRUEBA OBTENIDO MEDIANTE LA MACRO DE LA FIGURA 5 DESPUÉS DE REALIZAR 1000 ITERACIONES



El tratamiento riguroso de la precisión obtenida en las estimaciones por simulación es un tema complejo que desborda las pretensiones de este artículo. Sin embargo, una forma rápida de obtener información sobre este aspecto consiste en repetir varias veces la simulación, analizando la variabilidad que presentan los resultados obtenidos.

Repetiendo 10 veces el proceso de simulación anterior, se han obtenido los valores:

	$\alpha = 0.05$	$\alpha = 0.01$
	0.407694	0.501664
	0.406042	0.492733
	0.411207	0.505006
	0.407567	0.499914
	0.397649	0.487519
	0.411169	0.473205
	0.410448	0.517766
	0.403493	0.478370
	0.402846	0.481353
	0.419733	0.496663
Media (\bar{x}):	0.40778	0.49342
Desviación tipo (s):	0.00598	0.01360

Y a partir de estos valores podemos calcular los intervalos de confianza del 95%, que resultan ser: [0.40351; 0.41206] para $\alpha = 0.05$ y [0.48369; 0.50315] para $\alpha=0.01$. En las tablas se encuentran 0.410 y 0.490 respectivamente.

La revista "Teaching Statistics", dedicada a los aspectos docentes de la estadística, incluye artículos dedicados a estos temas, como el de Caster y Mougeot (1995) que explica cómo utilizar técnicas de simulación para comparar de forma muy didáctica las características que presentan varios tipos de estimadores de un parámetro dado, según sea la distribución de la población.

3. APLICACIÓN DE LA SIMULACIÓN A LA RESOLUCIÓN DE PROBLEMAS DE CÁLCULO DE PROBABILIDADES

Los capítulos de cálculo de probabilidades que aparecen en bastantes cursos de estadística de carreras universitarias, incluyen problemas que no sólo requieren tener los conceptos muy claros sino también una cierta práctica e incluso en muchas ocasiones la aparición de una "idea feliz".

Son problemas que pueden verse claros cuando otra persona explica, pero que no es fácil que a personas con poco entrenamiento se le ocurran. Creemos que no son pocos los alumnos que se ven desbordados por estos temas y que inician de esta forma una mala relación y un distanciamiento hacia la estadística y el cálculo de probabilidades, que empiezan a considerar como asignaturas que no les van.

No sólo los estudiantes, Morgan et al. (1991) presenta un caso de problema sencillo en su formulación (el conocido problema de las tres puertas), que intentaron resolver sin éxito muchas personas, entre ellas muchas procedentes del mundo universitario en los Estados Unidos (estudiantes, doctorados, profesores).

La simulación puede ser una excelente y accesible forma de abordar muchos problemas de cálculo de probabilidades y de llegar a una aproximación suficientemente buena del resultado buscado. Aunque como dice Simon (1994) no debe considerarse la simulación como un atajo a los resultados analíticos, sino como un camino diferente, que utiliza distintos 'inputs', una metodología distinta, y un resultado también cualitativamente distinto aunque no por ello necesariamente menos útil.

A continuación se presentan dos ejemplos de resolución de problemas de cálculo de probabilidades aplicando técnicas de simulación.

3.1. El Problema de los Sobres y las Cartas

Se trata de un problema clásico que presenta un resultado sorprendente (mayor del que intuitivamente parece) y que puede expresarse elegantemente mediante una expresión en la que interviene el número e . El problema es el siguiente: Se dispone de n sobres, cada uno de ellos con la dirección de una persona distinta y n cartas dirigidas a cada una de las personas que figuran en los sobres. Si las cartas se colocan al azar en los sobres, ¿cuál es la probabilidad de que como mínimo una carta se coloque en el sobre que le corresponde?. (La solución es: $1-1/e$, el 63.21%) (4).

La resolución mediante simulación puede abordarse de la siguiente forma:

Damos un valor a n , por ejemplo $n = 500$. En una columna escribimos los números del 1 al 500 ordenados y en otra los mismos números pero colocados al azar. Una columna con datos ordenados puede reordenarse al azar generando valores aleatorios de una distribución continua y ordenando esos valores arrastrando en esta ordenación a los números que se desea reordenar.

La resta de las columnas ordenada y reordenada al azar permite detectar si algún número coincide en la posición de ambas columnas. Una forma de realizarlo consiste en elevar al cuadrado las diferencias y buscar el mínimo de estas (MINITAB lo hace directamente). Si dicho mínimo es cero es que existe alguna coincidencia y si es distinto no la hay. Repitiendo este proceso k veces, la proporción de veces que haya coincidencia será una aproximación a la probabilidad buscada.

Utilizando MINITAB se puede poner en marcha la simulación utilizando una *macro* como la que se indica en la figura 7.

Figura 7

MACRO PARA RESOLVER POR SIMULACIÓN EL PROBLEMA DE LOS SOBRES Y LAS CARTAS. TERMINADO EL PROCESO, EL NÚMERO DE CEROS DE LA COLUMNA C5 INDICA EL NÚMERO DE VECES QUE HA HABIDO COINCIDENCIA

```
# Previamente hay que definir k1=1
SET C1                # Introduce valores en la columna 1 ...
(1:500)                # ... de 1 al 500 ordenados
END
RANDOM 500 C2          # Coloca 500 valores N(0,1) en columna 2
SORT C2 C1 C3 C4      # Ordena los valores de c2 arrastrando c1
LET K2=MIN((C1-C4)**2) # Realiza la operación indicada
LET C5(K1)=K2         # Acumula el mínimo obtenido
LET K1=K1+1           # Suma 1 al contador
PRINT K1              # Aparece en pantalla el núm. de iteración
END
```

(4) En Paulos (1993) puede leerse un comentario sobre este problema en un tono de divulgación general. En DeGroot (1988) o, con más detalle, en Feller (1968) puede encontrarse su resolución analítica.

Repitiendo el proceso 1000 veces se ha obtenido una proporción de coincidencias del 64.1 %. Al igual que se hizo con los valores de la distribución de referencia para el test de Kolmogorov, se puede evaluar la precisión de valor obtenido repitiendo varias veces el proceso de simulación. Realizando 10 repeticiones, hemos obtenido los valores:

64.1; 62.0; 66.4; 64.0; 61.0; 65.4; 62.0; 59.0; 61.2; 64.0

A partir de los cuales puede calcularse un intervalo de confianza del 95 % para el valor buscado, obteniéndose [61.30; 64.52].

Otro aspecto que puede resultar interesante introducir es el de la convergencia. Para ello es útil construir un gráfico como el de la figura 8, en el que se representa la evolución del porcentaje de coincidencias al ir aumentando el número de simulaciones. También resulta ilustrativo el gráfico de la figura 8 bis, tanto para poner de manifiesto la convergencia como la evolución de la precisión obtenida.

Figura 8

EVOLUCIÓN DEL PORCENTAJE ACUMULADO DE CASOS EN QUE SE PRESENTAN COINCIDENCIAS, AL IR AUMENTANDO EL NÚMERO DE SIMULACIONES

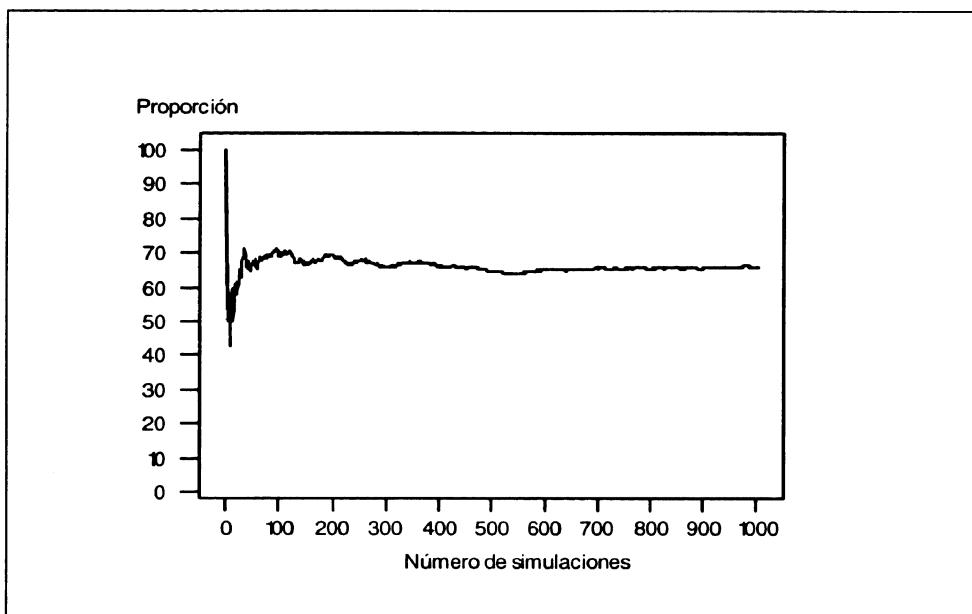
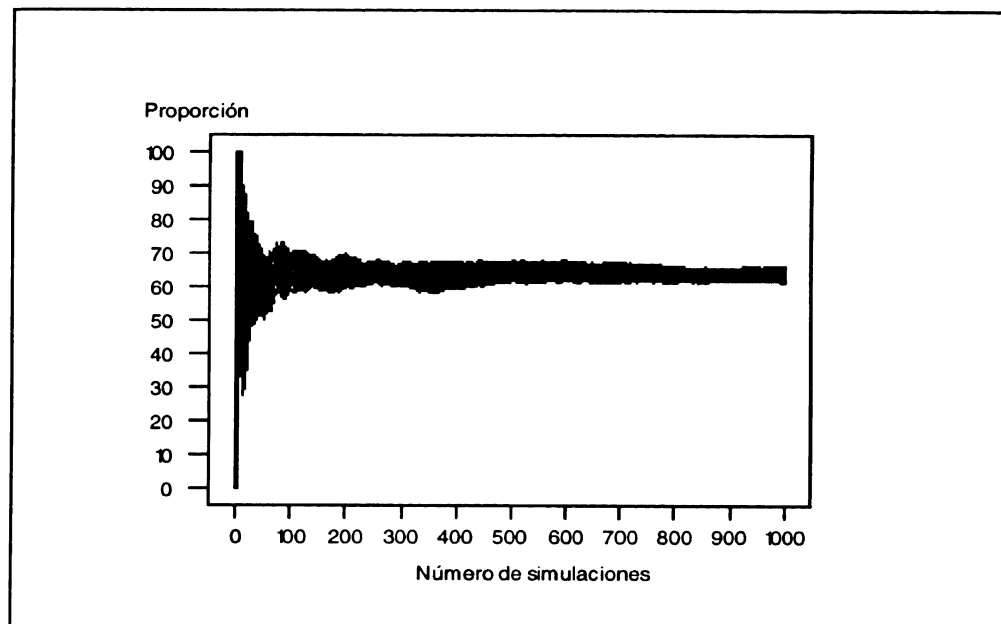


Figura 8 bis

ZONA DEFINIDA POR EL PORCENTAJE DE COINCIDENCIAS AL SUPERPONER LAS EVOLUCIONES OBTENIDAS EN LAS 10 REPETICIONES DEL PROCESO DE SIMULACIÓN



3.2. El Problema de la Colección de Cromos

Un pastelito, de esos que a los niños les gusta tomar en la merienda, incluye un cromo en cada bolsa. Si la colección consta de 60 cromos y todos tienen la misma probabilidad de aparecer, ¿cuántos pastelitos habrá que comprar para completar la colección?. Naturalmente no se trata de un valor exacto sino que interesará conocer su valor medio, la variabilidad que presenta y la forma de su distribución.

Este problema, que seguramente muchos nos hemos planteado en alguna ocasión, no tiene una solución analítica fácil pero sí puede abordarse sin grandes complicaciones utilizando técnicas de simulación.

Utilizando MINITAB el procedimiento es el siguiente: Se genera una columna con 60 valores "missing" y se van generando números aleatorios de una distribución equiprobable definida en el intervalo de 1 a 60. Cada número aleatorio obtenido (número de cromo que aparece en el pastelito) se coloca, en la posición que indica su número, en la columna que inicialmente sólo contenía valores "missing".

El proceso acaba cuando ya no quedan valores "missing" y el número de pastelitos comprados es igual a la cantidad de números aleatorios generados.

Por ejemplo, colocamos los 60 valores "missing" en la columna C1, generamos un número aleatorio de la distribución indicada y resulta ser el 23. En este caso lo colocaremos en la posición 23 de C1 haciendo $C1(23)=23$ y ya sólo quedarán 59 valores "missing". Posteriormente si el número 23 vuelve a salir se repetirá la misma operación, pero el número de valores "missing" no habrá cambiado, de forma que seguiremos teniendo el mismo número de cromos distintos que antes.

La macro que realiza este proceso se indica en la figura 9.

Figura 9

MACROS, Y PROCEDIMIENTO A SEGUIR, PARA SIMULAR UNA COLECCIÓN DE CROMOS

```
# Macro 'SOLOINI' para inicializar el proceso:
ERASE C1-C5      # Borra el contenido de las columnas de c1 a c5
SET C1           # Coloca en c1 ...
('*)'60          # ... 60 valores missing (*)
END              # Fin de la entrada de datos
LET K1=1         # Asigna el valor 1 a la constante k1
NOECHO
END

# Macro 'SOLO1' para simular la obtención de un cromo:
RANDOM 1 C2;      # Sólo un cromo por sobre
INTEG 1 60.
LET K2=C2(1)     # Llama k2 al número de cromo
LET C1(K2)=K2    # Sustituye el '*' por el número de cromo
NMIS C1 K1       # Llama k1 al número de cromos que faltan
LET C3(1)=K1     # Coloca el n°. de cromos que faltan en 1ª fila de
c3
STACK C3 C4 C4   # Acumula la cantidad de cromos que faltan en C4
END

# Macro 'SOLO' para simular una colección:
EXECUTE 'SOLO1' K1 # Va obteniendo cromos mientras faltan
END

Procedimiento de simulación: En primer lugar realizar EXECUTE
'SOLOINI' para inicializar el proceso; a continuación hacer EXECUTE
'SOLO' 800 para simular la colección de cromos. El valor 800 es el
número de máximo de cromos que se supone necesario para completar
la colección; si se completa antes no se ejecutará la macro 'SOLO1'
ya que k1=0.
```

Pero lo que interesa es conocer la distribución que presenta el número de pastelitos necesario hasta completar la colección. Esto puede realizarse mediante la macro de la figura 10 que contiene básicamente la ya incluida en la figura 9 más las instrucciones de acumulación del número de cromos necesarios para completar cada colección y una macro de inicialización cada vez que se completa una colección.

Figura 10

MACRO PARA SIMULAR COLECCIONES DE CROMOS (TANTAS COMO VECES SE EJECUTE)

```
EXECUTE 'SOLO' 800 # Simula una colección
N C4 K3           # Núm. cromos necesario completar colección en
K3
LET C5(1)=K3      # Coloca en la 1a. posición de c5 el valor de k3
STACK C5 C6 C6    # Va acumulando en c6 los valores de k3
EXECUTE 'SOLOINI' # Prepara la simulación de una nueva colección
END
```

La figura 11 representa el histograma de los valores obtenidos tras simular 1000 colecciones. El valor medio es 281.77, la mediana resulta ser 270.5 y el percentil del 90 % es 379 (5).

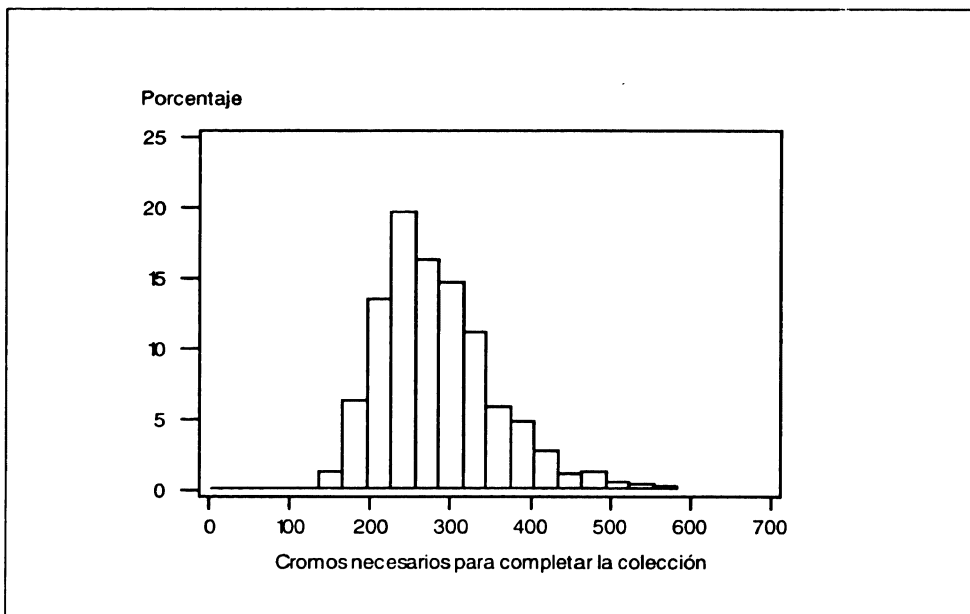
(5) La expresión analítica que da la probabilidad de completar una colección de k cromos teniendo N , suponiendo que todos los cromos aparecen con la misma probabilidad y que $N \geq k$ es (Rohatgi, 1976):

$$1 - \binom{k}{1} \left(\frac{k-1}{k} \right)^N + \binom{k}{2} \left(\frac{k-2}{k} \right)^N - \binom{k}{3} \left(\frac{k-3}{k} \right)^N + \dots + (-1)^{k-1} \binom{k}{k-1} \left(\frac{1}{k} \right)^N$$

Los valores que dan 0.5 y 0.9 (mediana y percentil del 90%) son 267 y 378 respectivamente.

Figura 11

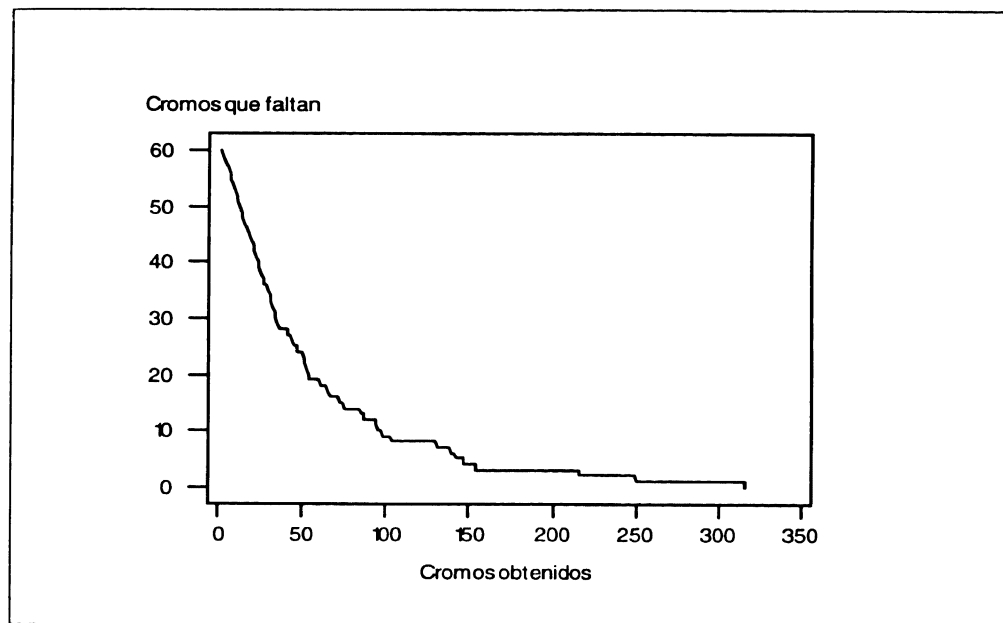
HISTOGRAMA QUE REPRESENTA LOS VALORES DEL NÚMERO DE CROMOS QUE SE DEBEN CONSEGUIR PARA COMPLETAR UNA COLECCIÓN DE 60, A PARTIR DE LA SIMULACIÓN DE 1000 COLECCIONES



La figura 12 presenta la evolución de una colección tipo que se completa con 317 cromos. Esta figura constata de forma clara algo que hemos vivido todos los que alguna vez hemos realizado colecciones de este tipo. Existe una primera etapa (aproximadamente hasta completar la mitad de la colección) que podríamos denominar “de ilusión”, ya que una alta proporción de los cromos que se consiguen son nuevos (“faltan”) y, por tanto, parece que completar la colección será fácil. A continuación se entra en una segunda etapa en la que aumenta considerablemente el número de cromos repetidos y que podríamos denominar “de expectación”, ya que ante cada cromo nuevo existe la expectación de saber si será repetido o no. Finalmente, cuando faltan 5 ó 10 cromos para acabar, se entra inevitablemente en la fase que podríamos denominar “de desesperación”, ya que el número de repetidos crece enormemente y cuesta mucho que aparezcan precisamente esos pocos cromos que faltan.

Figura 12

EVOLUCIÓN DEL NÚMERO DE CROMOS QUE FALTAN FRENTE AL NÚMERO DE CROMOS CONSEGUIDOS, PARA UNA COLECCIÓN TIPO QUE SE COMPLETA CON 317 CROMOS



¿Y qué pasa si se cambian cromos?

En este caso el tema se complica, especialmente si intervienen más de 2 colecciones, ya que es necesario establecer unas reglas del juego (quien cambia con quien y con qué prioridad). Hemos realizado la simulación en el caso de que realicen la colección 2 amigos y se regalen uno a otro los cromos repetidos (sin necesidad de que el otro le dé un cromos a cambio, pero considerando que cuando el compañero completa su colección ya no consigue más cromos y, por tanto, ya no puede seguir cambiando). El resultado obtenido de los cromos necesarios para completar 1000 colecciones realizadas de esta forma se indica en la figura 13. En la figura 14 se compara la distribución de cromos necesarios para completar la colección sin realizar cambios y con cambios de acuerdo con las condiciones indicadas.

Figura 13

DISTRIBUCIÓN DEL NÚMERO DE CROMOS QUE ES NECESARIO CONSEGUIR PARA COMPLETAR UNA COLECCIÓN DE 60 (A PARTIR DE LA SIMULACIÓN DE 1000 COLECCIONES), CONSIDERANDO QUE LOS REPETIDOS SE CAMBIAN CON OTRA PERSONA

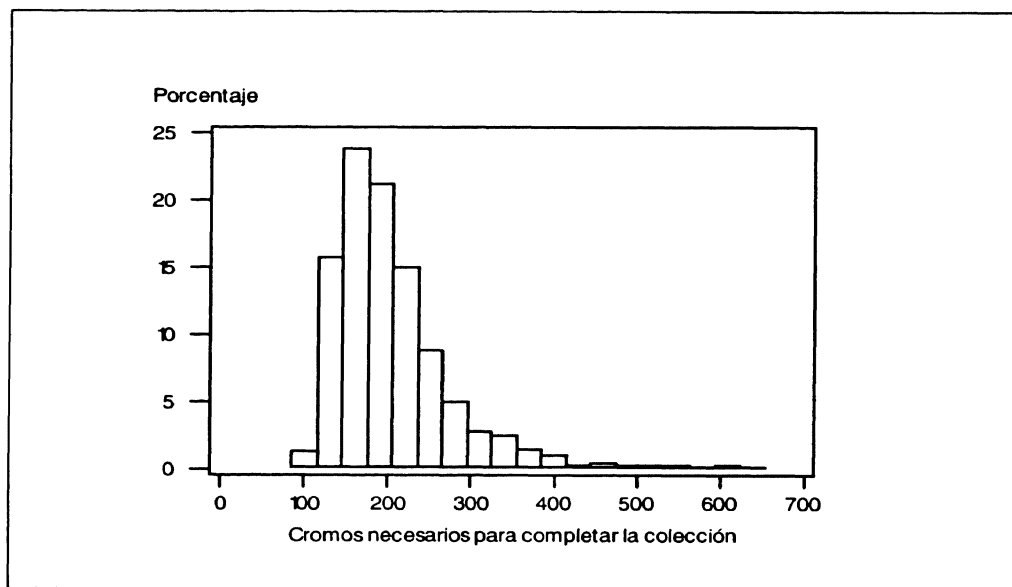
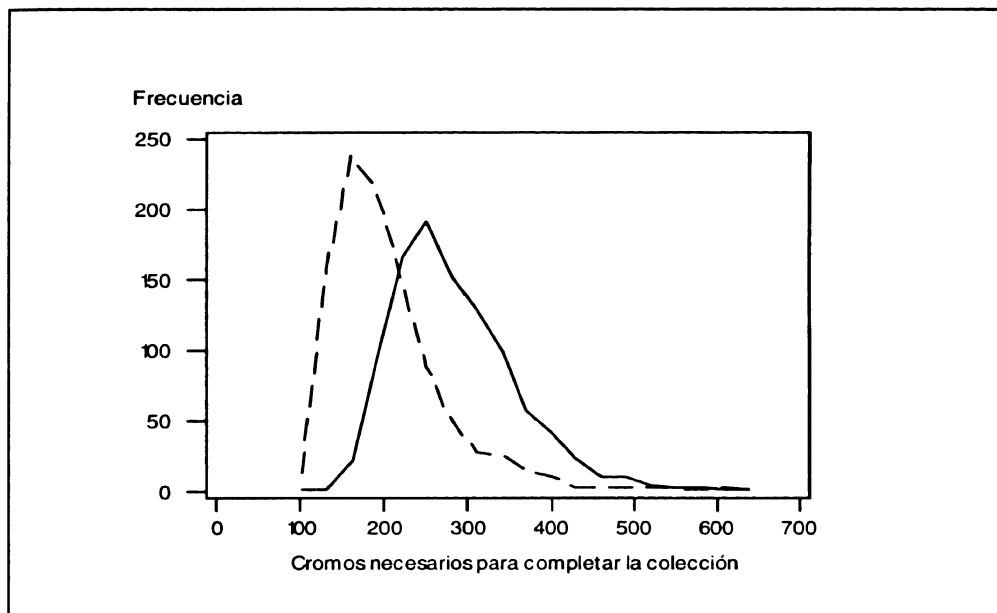


Figura 14

COMPARACIÓN DE LOS NÚMEROS DE CROMOS NECESARIOS PARA COMPLETAR LA COLECCIÓN DE 60 CAMBIANDO LOS REPETIDOS CON OTRA PERSONA (LÍNEA DISCONTINUA) O SIN CAMBIOS (LÍNEA CONTINUA)



¿Y si la colección consta de más cromos? ¿y si se pueden solicitar los x últimos que faltan? Estas son variantes que pueden contemplarse sin ninguna dificultad.

Naturalmente, existen multitud de problemas que pueden ser abordados a través de la simulación. Simon (1994) presenta seis interesantes problemas de cálculo de probabilidades resueltos de esta forma. Los utiliza para demostrar que aunque la formulación matemática pueda ayudar a algunas personas a mejorar su comprensión del problema (o en general, del mundo real), en muchos otros casos los métodos de muestreo (simulación y remuestreo entre ellos) pueden ser más efectivos que los métodos tradicionales. En Graham (1995) se explica cómo utilizar una hoja de cálculo para resolver problemas de cálculo de probabilidades por simulación, y en Konold (1995) comenta como la simulación puede ser útil para enseñar a resolver este tipo de problemas.

4. ASPECTOS RELACIONADOS CON LA VARIABILIDAD Y EL DISEÑO Y LA FABRICACIÓN DE PRODUCTOS

La simulación también puede ser utilizada para ilustrar conceptos relacionados con la aplicación de la estadística a campos específicos. A continuación se presentan 2 ejemplos que ponen de manifiesto fenómenos relativos a la influencia de la variabilidad en la determinación de los parámetros de diseño de los productos y a cómo comportarse frente a la variabilidad inevitable de los procesos de producción.

4.1. Determinación de los parámetros de diseño de un producto: ¿Cuánto debe valer una resistencia?

Seguramente una de las leyes físicas más conocidas por los estudiantes de carreras técnicas es la Ley de Ohm, que dice que la intensidad que circula por un circuito es directamente proporcional al voltaje que se le aplica e inversamente proporcional a su resistencia, de acuerdo con la fórmula:

$$I = \frac{V}{R}$$

Supongamos que tenemos que fabricar circuitos que consten únicamente de una fuente de alimentación y una resistencia. Supongamos asimismo que las fuentes de alimentación tienen un valor constante igual a 100 voltios pero que las resistencias presentan una cierta variabilidad en torno al valor nominal elegido siguiendo una distribución normal con desviación tipo igual a 4 Ω . Si deseamos que la esperanza matemática de la intensidad que circule por cada uno de los circuitos sea de 5 amperios, ¿cuál debe ser el valor nominal de las resistencias?

La solución analítica de este problema parece trivial pero no lo es. La respuesta no es 20 Ω ya que, en general:

$$E\left(\frac{V}{R}\right) \neq \frac{E(V)}{E(R)}$$

El problema puede ponerse de manifiesto por simulación generando valores de una distribución normal con media igual al valor nominal de la resistencia y la desviación tipo indicada. El cociente de dividir 100 (el valor de la fuente de alimentación) por cada uno de los números aleatorios obtenidos (valores de las resistencias) proporcionará un conjunto de valores de la intensidad de los cuales se puede calcular la media.

Utilizando MINITAB puede realizarse este proceso de forma iterativa barriendo un determinado rango de variación de valores nominales de la resistencia. En la figura 15 se incluye la macro que obtiene la media de las intensidades calculadas a partir de 5000 valores aleatorios de la resistencia obtenidos de una distribución $N(\mu; \sigma)$ con valores de μ que van desde 18 a 22 Ω con incrementos de 0.1 Ω y una $\sigma = 4$ Ω . En la figura 14 puede observarse el gráfico obtenido, que relaciona el valor no-

minal de la resistencia con la media de las intensidades obtenidas. El valor nominal óptimo de R para conseguir que $E(I) = 5$ A, resulta estar en torno a 20.8Ω .

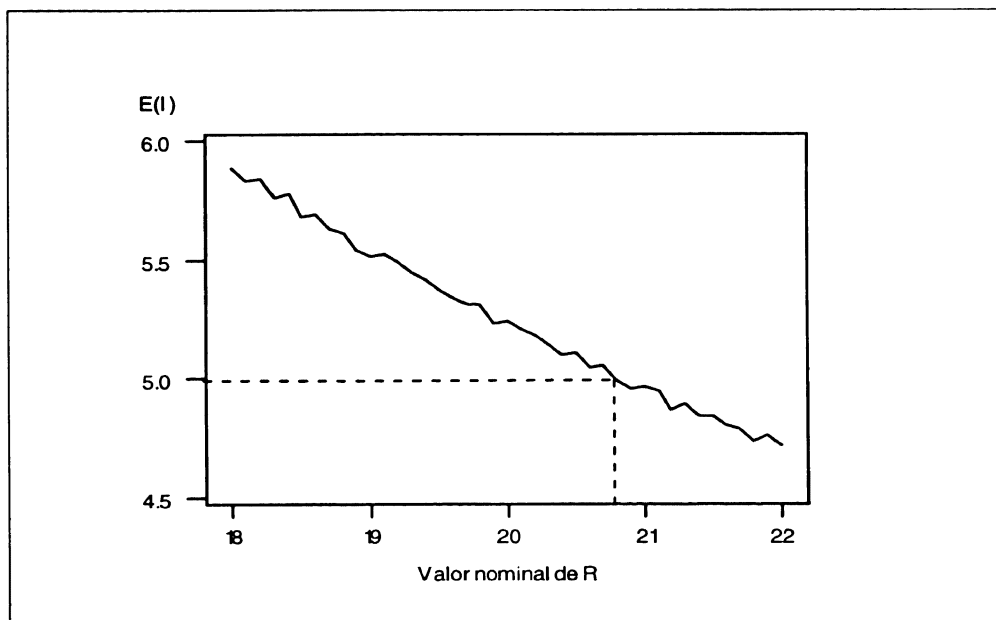
Figura 15

MACRO PARA DETERMINAR LA ESPERANZA MATEMÁTICA DE LA INTENSIDAD QUE CIRCULARÁ POR LOS CIRCUITOS DEL EJEMPLO, EN FUNCIÓN DEL VALOR NOMINAL DE LA RESISTENCIA

```
# Si se desea "barrer" el rango de valores de R de 18 a 22 W se
# deberá definir previamente k1=18 y ejecutar la macro 41 veces.
RANDOM 5000 C1;      # Genera 5000 números aleatorios
NORMAL K1 4.         # ... de una dist. normal con m=k1 y s=4
LET C2=100/C1        # Coloca en c2 100/valores generados
LET C3(1)=MEAN (C2)  # Coloca en c3 la media de los cocientes
LET C4(1)=K1         # Coloca en c4 el valor de m
STACK C5 C3 C5       # Acumula en c5 las medias de los cocientes
STACK C6 C4 C6       # Acumula en c4 los valores de m
LET K1=K1+0.1        # Suma 0.1 a m para repetir el proceso
PRINT K1
END
```

Figura 16

GRÁFICO QUE REPRESENTA LA ESPERANZA MATEMÁTICA DE LA INTENSIDAD QUE CIRCULA POR EL CIRCUITO (EN ORDENADAS) EN FUNCIÓN DEL VALOR NOMINAL DE LA RESISTENCIA



4.2. Aumento de la Variabilidad por sobreajuste

Uno de los conceptos importantes del control de calidad es el que se refiere a la capacidad de una máquina.

La capacidad de una máquina es una medida de la variabilidad que presenta en su producción cuando sólo influyen causas de variabilidad aleatorias (forman parte intrínseca de la máquina y son inevitables). Es importante conocer esta variabilidad para constatar que es coherente con las tolerancias del producto (en este caso se dice que la máquina es capaz) y para poder diferenciar entre desviaciones respecto al valor nominal que pueden ser debidas a las causas aleatorias, de aquellas otras que por las características que presentan deben ser atribuibles a causas de variación asignables y que, por tanto, conviene identificar y eliminar (por ejemplo, la máquina se ha descentrado).

Un error que se produce con cierta frecuencia es el de sobreajustar una máquina pensando que de esta forma se va a disminuir su variabilidad. Sobreajustar es regular de forma sistemática independientemente de que la máquina esté descentrada o no. Por ejemplo, un proceso de envasado de sacos de pienso de 20 kg presenta una variabilidad intrínseca de ± 300 gr, por tanto, si un saco pesa 20,150 kg no hay razón para pensar que el proceso esté descentrado. Sin embargo, si no se es consciente de la variabilidad intrínseca, el operario que observa esta desviación seguramente regulará la máquina para que el peso objetivo sea 150 gr menor. Esta forma de proceder no sólo no disminuye la variabilidad sino que la aumenta.

Una forma de verlo es simulando una máquina centrada sujeta a las causas aleatorias de variabilidad y una máquina también centrada pero sobreajustada después de cada unidad. Para ello se simularon 1000 datos de una distribución $N(20; 0.1)$ y a continuación otros 1000 en la que el primer valor pertenece igualmente a una $N(20; 0.1)$ pero para los valores sucesivos la media de la distribución va cambiando ajustándose siempre en la diferencia entre el último valor obtenido y el valor objetivo, de acuerdo con el siguiente esquema:

Valor	Distribución a la que pertenece
x_1	$N(\mu_1=20; \sigma=0.1)$
x_2	$N[\mu_2=\mu_1+(20-x_1); \sigma=0.1]$
x_3	$N[\mu_3=\mu_2+(20-x_2); \sigma=0.1]$
\vdots	\vdots
\vdots	\vdots
x_i	$N[\mu_i=\mu_{i-1}+(20-x_{i-1}); \sigma=0.1]$
\vdots	\vdots
\vdots	\vdots
x_{1000}	$N[\mu_{1000}=\mu_{999}+(20-x_{999}); \sigma=0.1]$

La macro que permite realizar esta simulación con MINITAB se indica en la figura 17. La comparación de la dispersión que presentan los dos conjuntos de datos (ajustando y sin ajustar) se compara en la figura 18.

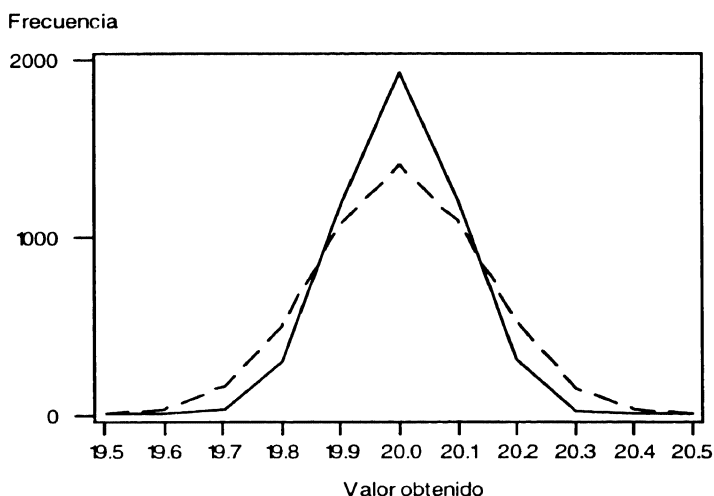
Figura 17

MACRO PARA SIMULAR LA OBTENCIÓN DE VALORES DE UNA DISTRIBUCIÓN QUE EMPIEZA SIENDO $N(20; 0.1)$, PERO QUE VA CAMBIANDO LA MEDIA DESPUÉS DE CADA OBSERVACIÓN EN UN VALOR IGUAL AL VALOR NOMINAL (20) MENOS EL ÚLTIMO VALOR OBTENIDO

```
# Previamente hay que definir: k1=1, k2=20
RANDOM 1 C1;           # Genera un valor ...
NORMAL K2 0.1.         # ... de una distribución N(k2, 0.1)
LET C2(K1)=C1(1)       # Va acumulando en c2 los valores obtenidos
LET K2=(20-C1(1))+K2    # Calcula la nueva media
LET K1=K1+1
PRINT K1               # Sirve para saber por donde va
END
```

Figura 18

POLÍGONOS DE FRECUENCIAS EN LOS QUE SE COMPARA LA VARIABILIDAD DE UNA DISTRIBUCIÓN NORMAL CON $\mu=20$ Y $\sigma=0.1$ (LÍNEA CONTÍNUA) CON LA QUE PRESENTA SI DESPUÉS DE CADA OBSERVACIÓN SE AJUSTA LA MEDIA (LÍNEA A TRAZOS)



Puede observarse que la dispersión de los datos que han sufrido un proceso de sobreajuste es mayor que la correspondiente a los que no se han ajustado. Exactamente, en el caso de un sobreajuste como el que aquí se ha realizado, la varianza será el doble ya que cada resultado obtenido responde a la expresión: $20 - y_{i-1} + y_i$, con $y_i \sim N(0; \sigma)$.

La conclusión es clara: no debe actuarse sobre un proceso si no se tienen razones fundadas para pensar que existen causas asignables de variabilidad. Si se actúa cuándo sólo intervienen causas aleatorias, la variabilidad no sólo no disminuye sino que aumenta. A partir de aquí está clara la necesidad de los estudios de capacidad y la conveniencia del control estadístico de procesos.

5. CONCLUSIONES

La difusión del ordenador personal y de los paquetes de software estadístico, permiten utilizar la simulación como una herramienta importante en la enseñanza de la estadística y el cálculo de probabilidades, ya que:

Permite constatar y visualizar importantes leyes, como la distribución de los estadísticos muestrales.

Ayuda a asimilar conceptos importantes como el de contraste de hipótesis y el funcionamiento de tests estadísticos como el de Kolmogorov.

Permite estimar los resultados cuando su cálculo es poco intuitivo, o cuando la resolución analítica sería muy complicada, máxime cuando el resultado no necesita de una gran precisión sino de una aproximación razonable.

Evita que los estudiantes se atasquen y se desanimen ante la resolución de problemas complicados de cálculo de probabilidades.

Permite comprobar los resultados obtenidos analíticamente.

El campo de aplicación de la simulación, en el contexto de la enseñanza de la estadística, es muy amplio y en este artículo se han presentado algunos ejemplos para poner de manifiesto estas posibilidades. Por otra parte, la constante evolución de la tecnología informática y de los paquetes de software hace suponer que la generación de números aleatorios y la realización de simulaciones se podrá realizar cada vez con mayor rapidez y facilidad.

AGRADECIMIENTOS

Los autores desean poner de manifiesto su agradecimiento a los profesores Albert Prat y Lourdes Pozueta por sus comentarios y sugerencias en la realización de este trabajo. Asimismo, desean agradecer a un evaluador anónimo de la revista el minucioso estudio que realizó del primer original y sus muchas y acertadas sugerencias, las cuales permitieron mejorarlo sustancialmente.

REFERENCIA

- Box, G.E.P.; Hunter, W.G. y Hunter, J.S.(1988): "Estadística para Investigadores". Reverté.
- CASTER, L.; MOUGEOT, M. (1995) "Using Computers to teach theoretical Statistics" *Teaching Statistics*, 17, 44-46.
- DEGROOT, M.H. (1988): "Probabilidad y Estadística". Addison-Wesley.
- DEVROYE, L. (1986): "Non-Uniform Random Variate Generation". New York: Springer-Verlag.
- FELLER, W. (1968): "An Introduction to Probability Theory and Its Applications". New York: John Wiley
- GRAHAM, A. (1995): "Intuition and Spreadsheet". *Teaching Statistics*, 17
- GINEBRA, J., et al. (1995): "Estadística. Cuaderno de Prácticas". Centre de Publicacions d'Abast. E.T.S. Ingenieros Industriales de Barcelona.
- KONOLD, K. (1995): "Confessions of a Coin Flipper and Would-Be Instructor" *The American Statistician*, 49, 203-209
- LEWIS, T.G. (1975): "Distribution Sampling for Computer Simulation". Lexington (Massachussets): Lexington Books.
- MORGAN, J.P., et al. (1991): "Let's Make a Deal: The Player's Dilema". *The American Statistician*, 45, 4, pp 284-289.
- NOVALES, A. (1989): "La enseñanza de la Estadística en las Facultades de Ciencias Económicas y Empresariales". *Estadística Española*, 31, pp. 331-355.
- PAULOS, J.A. (1993): "Más allá de los Números". Tusquets Editores. Colección Metatemas núm. 31.
- PEÑA, D. (1992): "Reflexiones sobre la enseñanza experimental de la estadística". *Estadística Española*, 34, pp. 469-490.

- PEÑA, D. (1991): "Estadística Modelos y métodos". Vol. 1: Fundamentos. Alianza Editorial.
- PEÑA, D.; PRAT, A.; Y ROMERO, R. (1990): "La Enseñanza de la Estadística en las Escuelas Técnicas". *Estadística Española*, 32, pp. 147-200.
- PRAT, A.; TORT-MARTORELL, X.; GRIMA, P.; POZUETA, L. (1994): "Métodos Estadísticos. Control y Mejora de la Calidad". UPC.
- ROHATGI, V. K. (1976): "An Introduction to Probability Theory and Mathematical Statistics". Wiley.
- SANTISTEBAN, C. (1989): "La Estadística en Psicología". *Estadística Española*, 32, pp. 461-475.
- SIMON, J.L. (1994): "What Some Puzzling Problems Teach About the Theory of Simulation and the Use of Resampling". *The American Statistician*, 48, 4, pp 290-293.
- WATTS, D.G. (1991): "Why is Introductory Statistics Difficult to Learn? And What Can We Do to Make it Easier?". *The American Statistician*, 45, 4, pp 290-291.

SUMMARY

This paper presents some ideas on how to exploit the simulation chances offered by statistical packages in the context of statistical teaching. Some practical examples are presented on topics such as verification of statistical laws, resolution of probability problems and some aspects related with variability in production processes.

The selected examples are solved, as an example, with the aid of MINITAB statistical package.

Keywords: Simulation, Sample Statistics Distribution, Probability, Statistical Packages

AMS Classification: 65C05, 68020, 6201