

Aleatorio o No Aleatorio...

```
int getRandomNumber()  
{  
    return 4; // chosen by fair dice roll.  
             // guaranteed to be random.  
}
```

1. **Lourdes Hernández Hernández**, PA Universidad Politécnica de Puebla. lurhdes@yahoo.es, y
2. **Santiago Daniel Markwalder Benítez**, PA Universidad Realística de México. santiagodaniel_m@yahoo.com.mx

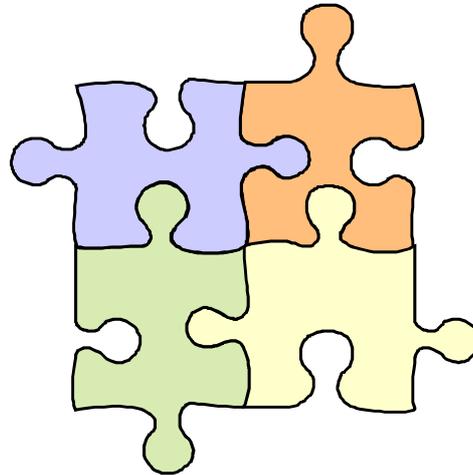
Resumen

Este trabajo intenta rescatar a una de las numerosas e interesantes herramientas estadísticas y tal vez poco comprendida como son los números aleatorios. Existen múltiples procedimientos que buscan hacer a la muestra totalmente representativa de la población, pero no es por sí misma suficiente para asegurar satisfacción, es importante considerar la estrategia de muestreo y el tipo de estimador aplicado, a fin de reproducir algunos totales poblacionales. Siempre permanece la incógnita acerca de cómo elegir dicha(s) muestra(s) (individuos), que tomamos en un determinado estrato o unidad (primera etapa). La única forma de hacerlo es seleccionando al azar o aleatoriamente. Cumplir tal condición asegura que la información obtenida a partir de la muestra es relevante a la población y no es simplemente una propiedad del método de selección. Elegir o escoger verdaderamente al azar o de manera aleatoria no es una tarea trivial o fácil de realizar. Varios artificios cuasi-aleatorios o pseudoaleatorios como lanzar monedas, barajar cartas, tirar dados, funciones deterministas, etcétera son algunas veces defendidos pero aun estos son responsables de introducir elementos no aleatorios dentro del proceso de selección.

Cualquier método confiable utiliza un cuadro o tabla de *números aleatorios*; se define un número aleatorio R como una variable aleatoria distribuida uniformemente entre cero y uno ($0 \leq r < 1$), es decir, cualquier valor en este rango tiene igual probabilidad de ocurrencia. Escritos en un cuadro en el cual cualquiera de las entradas posibles es igualmente probable que ocurra en cualquier posición dada.

Los números aleatorios tienen diferentes salientes entre las que destacan la forma de utilizarse, su complicada generación y su no menos sencilla aplicación, el objetivo de este trabajo se concentra principalmente en responder cuando y como emplear los números aleatorios, dentro de los trabajos de investigación dígame diseño de experimentos y estudios observacionales (encuesta, entrevista).

Introducción



Alguna vez le han planteado algún problema en esencia igual o parecido a los que a continuación se enuncian:

- a. Usted debe escoger 20 nombres y teléfonos al azar de un directorio telefónico a fin de aplicar una encuesta. ¿Que procedimiento seguiría? (considere que se tienen diferentes páginas, columnas y diferentes individuos en estas). Proporcione una forma alterna de escoger aleatoriamente estos números telefónicos.
- b. Se tiene que elegir de manera aleatoria 5 ratas de un grupo de 19 ratas de laboratorio para un experimento. Describa como procedería para asegurarse que la selección fue al azar o aleatoria. ¿De qué herramienta haría uso? en caso de así considerarlo necesario.
- c. Hay 32 estudiantes en una clase de Estadística repartidos en 4 filas de 8 estudiantes cada una. En cada clase se seleccionan a 4 de ellos aleatoriamente. ¿Cómo deberían ser seleccionados? A cuales estudiantes se deberían seleccionar durante las primeras 10 clases (identifique la forma exacta de seleccionarlos). ¿Alguien es elegido más de una vez?
- d. Es necesario aplicar de *manera aleatoria* o *al azar* una encuesta a 1000 consumidores en distintos lugares donde se consume un cierto producto en una ciudad importante. ¿Cómo lo haría? Garantizando que la elección fue de manera aleatoria.

Partiendo del hecho que debe existir una razón para recolectar datos, aunque la determinación acerca de que tipo de datos y como proceder puede no ser siempre clara. La mayoría de las consideraciones de que preguntas hacer en una encuesta, que mediciones realizar a un automóvil en una prueba de emisiones, y que químicos medir en la sangre al realizar un experimento, así como de que forma realizar estas mediciones, ciertamente es responsabilidad del investigador. Sin embargo, el número de pruebas y tanto como del diseño o arreglo del experimento dentro de lo posible debe ser determinado con un análisis estadístico.



A menudo tenemos datos que podríamos llamar *datos acumulados*, porque han sido recolectados esencialmente sin control tal vez ni idea (objetivo) de lo que se busca (cualquier coincidencia, mera casualidad!!!!).

Probablemente hubo un poco de planeación, como determinar donde establecer las locaciones de muestreo, los momentos o los tiempos de chequeo, pero no se tiene control sobre el clima, el número de eventos que ocurrirán, o como una contingencia no considerada afectará el patrón de comportamiento de la prueba. *Y aun más a menudo no hay control o información en la forma de recolectar los datos* –nos encontramos a merced de nuestro capricho o el de la gente designada para recoger la información-.

Alentadoramente, se ha *diseñado un experimento*. “Se corren pruebas para determinar el tiempo de llegada a la escuela, se debe escoger el número de repeticiones o réplicas. Se seleccionan los tiempos para iniciar de *manera aleatoria* de entre los tiempos posibles de inicio. Digamos que normalmente es posible iniciar en cualquier momento dentro de los primeros 10 minutos entre las 8 a las 8:09 a.m., es posible usar una “tabla” o cuadro de datos con *números aleatorios*” para elegir el tiempo de inicio del experimento”.

En el Cuadro 1 anexo, tenemos lo que se conoce como un arreglo de *números aleatorios*. Este consiste de dígitos, números del 0 al 9, colocados en tal forma que no importa donde se inicie, el número que aparezca puede ser cualquiera de los dígitos y con la misma posibilidad de aparecer con respecto a los 9 restantes. El siguiente dígito, sin importar en que dirección se elija ir, puede también ser cualquiera y con una posibilidad igual de aparecer. Así, si iniciamos en la primera columna de la fila 31 y observamos al primer arreglo de dígitos en la izquierda, encontraremos los números 40603 16152 83235.... y así sucesivamente (Ver cuadro 1).

El primer dígito es el número cuatro, así el primer día se iniciará la prueba a las 8:04 a.m. (nótese que el número 4 esta siendo usado como el número de minutos después de las 8:00 en punto, en este caso 4 minutos después de las 8:00 a.m. El segundo dígito de la lista es un 0, y este representa el número de minutos después de las 8 en punto en el segundo día, o 8:00 a.m. El tercer día iniciará a las 8:06 a.m., y así sucesivamente. Si seguimos este procedimiento estaremos seguros de que el tiempo de inicio no este sesgado o influenciado por alguien en su selección. Han sido escogidos de forma totalmente *aleatoria*, y asegurar la *aleatoriedad* al escoger nuestras muestras es vital.



Respondiendo a
nuestra pregunta planteada
en el inciso b. tenemos:

20174
74490
98704
22157
93291
45047
44919
16239
04506
32701
03645
62490
61009

Primero identificamos a las ratas de forma conveniente y las numeramos del 1 al 19; Nos ubicamos en un punto de partida en el cuadro de números aleatorios; una conveniente regla es la de iniciar donde nos quedamos la última vez que usamos esta herramienta. Una manera sencilla, pero no siempre la más eficiente, consiste en rechazar todos los números mayores que 20 (establecer que 00 cuenta como mayor que 20) y tomar los primeros cinco restantes, rechazar cualquier repetición. Suponga que nuestra selección se inicia al principio de la primera columna; leemos los pares de dígitos en las columnas 1 y 2, hacia abajo, este procedimiento da como resultado 20, 16, 4, 3, 1; las ratas con estos números son entonces nuestra muestra aleatoria.

89039
01723
27564
.....

Un segundo método más elaborado y eficiente es presentado a continuación. Divida cada par de dígitos entre 19 y tome el residuo (si el residuo es igual a cero para este propósito es considerado como 19); estos residuos son entonces los números aleatorios requeridos. Debido a que 19 no es un submúltiplo exacto de 10. Debido al tamaño del grupo, los pares de dígitos 1, ..., 19; 20, ..., 38; 39, ..., 57; 58, ..., 76; 77, ..., 95 todos dan residuos de 1, ..., 19, pero los pares 96, ..., 00 producen un arreglo incompleto de residuos 1, ..., 5 y por esta razón deben ser excluidos debido a que su uso alteraría las frecuencias de los residuos. Iniciando de arriba abajo a partir de la primera columna entonces seleccionaríamos los números 20, 74, ~~98~~, 22, 93, 45 y 44 que al ser divididos entre 19 (nótese la exclusión del 98) obtenemos aquellos individuos numerados con 1, 17, 3, [17], 7, 6. A menos que el muestreo sea requerido con reemplazo ignoraríamos el segundo 17.

En otro ejemplo, suponga que tenemos 40 latas de sopas deshidratadas para acampar, y que deseamos tomar una muestra de tamaño $n = 4$ para estudiar su condición.

Nuestro primer paso es numerar las cajas de 1 a 40 o apilarlas en algún orden de tal forma que puedan ser identificadas. En el Cuadro 1 adjunto de números aleatorios, los dígitos deben escogerse de a dos a la vez porque la población de tamaño $N = 40$ es un número de dos dígitos. Empezamos seleccionando arbitrariamente una página, una fila, y una columna de la tabla. Suponga que nuestra selección es la fila 46, y la columna 9. Leemos los pares de dígitos en las columnas:

31 76 78 48 74 27 37 37 99 96 66 81 79 52 19

Ignoramos los números mayores que 40 y también cualquier número repetido cuando aparezca una segunda vez, como el 37. Se continúa leyendo pares de dígitos hasta que cuatro unidades diferentes hayan sido seleccionadas.

31 27 37 19

Entonces se examinan los contenidos de las latas seleccionadas.

Otra forma de resolver este problema, para el caso específico de seleccionar aleatoriamente un miembro de una lista de N personas, se genera o escoge un número aleatorio r dividido entre 100, se puede demostrar que el elemento seleccionado será el que ocupe la posición dada por el resultado de multiplicar el número aleatorio r por el número de elementos de la lista N , sumarle uno y tomar la parte entera del resultado, es decir, la posición J seleccionada será la parte entera de la siguiente expresión: $J = [N r + 1]$. Si usáramos este procedimiento para los números

aleatorios escogidos (0.31, 0.27, 0.37 y 0.19) las latas a muestrear serían las correspondientes a los números $[40 \times 0.31 + 1, 40 \times 0.27 + 1, 40 \times 0.37 + 1 \text{ y } 40 \times 0.19 + 1]$ = $[13.4, 11.8, 15.8 \text{ y } 8.6]$ = $[13, 11, 15, \text{ y } 8]$.



Por supuesto, podemos ampliar el uso de los *números aleatorios* para incluir un mayor arreglo de dígitos. Por ejemplo, podemos asignar un número de 000 a 999 para cada estudiante de los 1000 estudiantes de nuestra población de estudiantes a fin de aplicarles un cuestionario. A continuación escogemos los números aleatorios tomando grupos de 3 dígitos, los cuales servirán como el número de estudiante a elegir. Así, si observamos en la primera columna de la Fila 32 del mismo arreglo de dígitos, y encontramos los números 409.

De esta forma el primer estudiante en ser entrevistado será el estudiante número 409. El siguiente estudiante será el 415, el tercero 358, el cuarto 569, y el siguiente 958. De esta forma seleccionaremos a nuestros estudiantes completamente *al azar*.

Para muestreos a gran escala o aplicaciones frecuentes, se recomienda usar la tabla con valores que vienen de una fuente más grande de número aleatorios llamados Un millón de números aleatorios con 100,000 Desviaciones Normales, por la Corporación Rand o un generador de número aleatorios de un computador, adecuadamente probado.

Regresando a nuestro diseño de experimentos, se debe notar que se pudieron tener problemas más complicados por resolver. Por ejemplo, un biólogo puede estudiar los niveles de azúcar en sangre de ratas macho y hembras de diferentes edades, en diferentes dietas, y a diferentes tiempos del día, usando un simple arreglo de roedores para contestar todas estas preguntas al mismo tiempo. El diseño de este experimento exige y va más allá de la estadística elemental, obviamente se debe estar consciente del alcance de las situaciones que pueden ser analizadas con técnicas estadísticas modernas (Diseños experimentales).

Finalmente, tenemos los datos tipo encuesta. Aquí entre los muchos puntos que son importantes, debemos considerar el hecho de que el cuestionario debe estar preparado apropiadamente y previamente validado, los individuos deben ser elegidos de manera adecuada y aleatoria. La información recabada deberá reflejar a la población en la que

estamos interesados entender; por lo tanto, es necesaria una cuidadosa y extensiva planeación.

“Nótese que si afirmamos que una muestra es representativa de la población quiere decir que conocemos las características de la población, entonces no tendría sentido hacer un estudio por muestreo”

En este momento tal vez el lector diga ¿Qué...? es posible ¡!!!, motivo por el cual vale la pena aclarar algunos puntos.

¿**Q**ué son los números aleatorios?

Son números que deben de cumplir los requisitos de ser equiprobables, es decir, que todo elemento tenga la misma probabilidad de ser elegido y que la elección de uno no dependa de la elección del otro.

¿**C**uál es su aplicación?

En la vida cotidiana se utilizan números aleatorios en situaciones como pueden ser los juegos de azar, Matemática Aplicada, Análisis Numérico, Criptografía (comprobar la cantidad de información filtrada por algoritmos criptográficos), estudios de Simulación de sistemas físicos o matemáticos (como el Método Monte Carlo para Cálculo Integral) para reproducir otra serie de fenómenos o variables aleatorias, para modelar y estudiar resultados de experimentos científicos en los sistemas electrónicos de telecomunicaciones. Los sistemas que usamos en los ordenadores actuales se basan en una semilla de base aleatoria (mover el ratón sin sentido mientras se realiza el cálculo)

Si escogemos números de nuestra mente se tenderá enfatizar a ciertos números y no incluir suficientemente a otros. ¿Cierto o Falso?

Para responder a la pregunta anterior. Escoja un número de un dígito al azar (0 a 9) y anótelo. Vuelva a hacer lo mismo tres veces más, cada vez olvidando o ignorando lo previamente realizado. De esta forma cada dígito es aleatorio y usted tendrá un arreglo de 4 dígitos aleatorios. Repita este proceso o procedimiento 25 veces más. Tendrá un total de 100 dígitos en grupos de 4.

- Cuente el número de veces que usted anotó cada dígito. Debería tener aproximadamente 10 de cada uno. ¿Qué tan cerca estuvo?
- Examine cada uno de los 25 grupos de 4. Determine cual grupo tiene 2 (o más) de los mismos dígitos (p.ej.: 8, 6, 1, 8 tiene 2 “ochos”). ¿Qué porcentaje de sus grupos tiene dígitos repetidos?

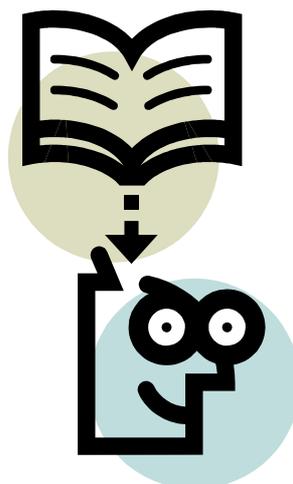
- c. Ahora examine el cuadro adjunto con números aleatorios. Observe a 25 arreglos de 4 dígitos. ¿Qué porcentaje de sus grupos tiene dígitos repetidos?
¿Qué comentarios puede hacer?

¿Cómo se generan?

El método usado para producir verdaderos números aleatorios es sumamente difícil y requiere de técnicas especiales, que no cualquiera posee. Un requisito es el de tener un comportamiento caótico, como la desintegración radiactiva o el ruido térmico en los transistores, impredecibles para cualquier tipo de aplicación práctica.

Se denomina **generador** de números aleatorios a cualquier procedimiento que produzca números entre cero y uno. Los números generados, además de distribuirse uniformemente, deben ser independientes, de tal forma que los fenómenos que reproduzcan no estén correlacionados entre sí. Se busca generar secuencias de números que no caigan dentro de una rutina, es decir, un ciclo de elementos no debe acabar repitiéndose.

Póngamelos para llevar...



Esta sección cubre la parte complementaria de proponer algunos problemas para que el lector sopesa las nuevas habilidades adquiridas.

- Para investigar irregularidades electorales, una muestra aleatoria simple de tamaño 60 es tomada de una lista de 1024 votantes inscritos en un barrio particular. ¿Como procedería a hacer su elección?, garantizando aleatoriedad
- Un estudio de nutrición es dirigido para determinar la calidad de la comida ingerida por niños en una escuela primaria. De un total de 1500 niños en escuelas primarias en una ciudad en particular, se debe seleccionar una muestra aleatoria simple de 80. ¿qué procedimiento emplearía? ¿por qué?
- Un guardabosque desea estimar la proporción de árboles en un bosque de tierras bajas que tienen infecciones de madera mojada. Estas infecciones pueden descubrirse por la presencia de una bacteria particular en el jugo extraído de los

agujeros taladrados. Se sabe que hay aproximadamente 5000 árboles en el bosque. Se decide muestrear 400 árboles, para realizarles las correspondientes pruebas. ¿De qué forma se llevaría a cabo el muestreo? ¿Garantiza ser aleatorio?

- Los errores que se presentan en las cuentas por cobrar son de importancia vital para los interventores. Un interventor que trabaja para una compañía de transportes desea estimar la proporción de casos en los que clientes reciben facturas defectuosas. Suponga que durante un periodo dado de tiempo, se archivan 2325 cuentas por cobrar en la oficina del interventor. Se elige muestrear a 500 de estas cuentas. Indique el procedimiento.
- Suponga que 588 granjas localizadas en un área particular constituyen una población y que el gasto de capital de su último año en maquinaria y en equipo de la hacienda es la característica a ser estudiada. Se escogen 60. Usando los números aleatorios correspondientes señale el método que emplearía para llevar a cabo semejante tarea.

Bibliografía



ESTUDIOS DE MUESTREO ¹Traducción del capítulo 16 "Sample Surveys" del libro "Statistical Concepts And Methods", Gouri K. Bhattacharyya and Richard A. Johnson. John Wiley and Sons, Inc. 1977. Traducción por parte de estudiantes del curso de "Estadística Matemática I" (semestre 02 de 2000) del Departamento de Ingeniería Industrial de la Facultad de Ingeniería de la Universidad de Antioquia, con la recopilación, corrección y adaptación del profesor Bernardo A. Calderón C. http://siona.udea.edu.co/~bcalderon/5_preguntamuestreo.html (Revisado en mayo del 2008)

Campbell, R. C. 1975. STATISTICS FOR BIOLOGISTS. Second Edition. Cambridge University Press. p. 5-6; 343-344. ISBN 0-521-09836-x

Un gaditano en Silicon Valley. Un millón de números aleatorios.
<http://folksonomy.com/un-millon-de-numeros-aleatorios/> (Revisado en mayo del 2008)

Ullman, Neil R. 1978. Elementary Statistics. An Applied Approach. John Wiley and Sons, Inc. ISBN 0-471-02105-9

Triola, Mario F. 2006. Estadística. Novena Edición Traducida al Español. Pearson Educación. México. p.24. ISBN 970-26-0519-9

Cochran, William G. 1992. Técnicas de Muestreo. Primera Edición. 9a Reimp. Traducida al Español. CECSA.México. p. 42-43. ISBN 968-26-0151-7.

Cuadro 1. Números Aleatorios. Sin importar el punto de inicio a partir del cual se elijan o la dirección en la cual se continué seleccionando valores. El modelo en que está basada la tabla de números aleatorios asegura que todos los dígitos simples tienen la misma probabilidad de ocurrencia de 1/10, que todos los pares de dígitos 00, 01, ..., 99 tienen una probabilidad de ocurrencia igual a 1/100, y así sucesivamente. Es posible ir hacia la derecha, izquierda, arriba, abajo y/o diagonalmente. Se presentan en grupos de cinco meramente por conveniencia. Aquellos marcados con gris, responden a ejemplos señalados en el cuerpo del texto.

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	44482	66558	37649	08882	90870	12462	41810	01806	02977
36792	26236	33266	66583	60881	97395	20461	36742	02852	50564
73944	04773	12032	51414	82384	38370	00249	80709	72605	67497
49563	12872	14063	93104	78483	72717	68714	18078	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	79207	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
40603	16152	83235	37361	98783	24838	39793	80954	76865	32713
40941	53585	69958	60916	71018	90561	84505	53980	64735	85140
73505	83472	55953	17957	11446	22618	34771	25777	27064	13526
39412	16013	11442	89320	11307	49396	39805	12249	57656	88686
57994	76748	54627	48511	78646	33287	35524	54522	08795	56273
07428	58863	96023	88936	51343	70958	96768	74317	27176	29600
35379	27922	28906	55013	26937	48174	04197	36074	65315	12537
10982	22807	10920	26299	23593	64629	57801	10437	43965	15344
90127	33341	77806	12446	15444	49244	47277	11346	15884	28131
63002	12990	23510	68774	48983	20481	59815	67248	17076	78910
40779	86382	48454	65269	91239	45989	45389	54847	77919	41105

43216	12608	18167	84631	94058	82458	15139	76856	86019	47928
96167	64375	74108	93643	09204	98855	59051	56492	11933	64958
70975	62693	35684	72607	23026	37004	32989	24843	01128	74658
85812	61876	23570	75754	29090	40264	80399	47254	40135	69911
61834	59199	15469	82285	84164	91333	90954	87186	31598	25942
91402	77227	79516	21007	58602	81418	87838	18443	76162	51146
58299	83880	20125	10794	37780	61705	18276	99041	78135	99661
40684	99948	33880	76413	63839	71371	32392	51812	48248	96419
75978	64298	08074	62055	73864	01926	78374	15741	74452	49954
34556	39861	88267	76068	62445	64361	78685	24246	27027	48239
65990	57048	25067	77571	77974	37634	81564	98608	37224	49848
16381	15069	25416	87875	90374	86203	29677	82543	37554	89179
52458	88880	78352	67913	09245	47773	51272	06976	99571	33365
33007	85607	92008	44897	24964	50559	79549	85658	96865	24186
38712	31512	08588	61490	72294	42862	87334	05866	66269	43158
58722	03678	19186	69602	34625	75958	56869	17907	81867	11535
26188	69497	51351	47799	20477	71786	52560	66827	79419	70886
12893	54048	07255	86149	99090	70958	50775	31768	52903	27645
33186	81346	85095	37282	85536	72661	32180	40229	19209	74939
79893	29448	88392	54211	61708	83452	61227	81690	42265	20310
48449	15102	44126	19438	23382	14985	37538	30120	82443	11152
94205	04259	68983	50561	06902	10269	22216	70210	60736	58772
38648	09278	81313	77400	41126	52614	93613	27263	99381	49500
04292	46028	75666	26954	34979	68381	45154	09314	81009	05114
17026	49737	85875	12139	59391	81830	30185	83095	78752	40899
48070	76848	02531	97737	10151	18169	31709	74842	85522	74092
30159	95450	83778	46115	99178	97718	98440	15076	21199	20492
12148	92231	31361	60650	54695	30035	22765	91386	70399	79270
73838	77067	24863	97576	01139	54219	02959	45696	98103	78867

De: RAND Corporation, Un Millón de Números Aleatorios con 100,000 Desviaciones Normales, publicadas en 1955 por la Free Press.