

APUNTES SOBRE TÉCNICAS DE MUESTREO DE WILLIAM COCHRAN

ISADORE NABI

<i>I. EL PAPEL DE LA TEORÍA DEL MUESTREO, EL MUESTREO PROBABILISTA Y EL USO DE LA DISTRIBUCIÓN NORMAL</i>	<i>2</i>
<i>II. EL SESGO Y SUS EFECTOS</i>	<i>5</i>
<i>III. ERROR CUADRÁTICO MEDIO</i>	<i>11</i>
<i>IV. MUESTREO ALEATORIO SIMPLE</i>	<i>12</i>
<i>IV.I. Generalidades</i>	<i>12</i>
<i>IV.II. Selección de una Muestra Aleatoria Simple</i>	<i>14</i>

I. EL PAPEL DE LA TEORÍA DEL MUESTREO, EL MUESTREO PROBABILISTA Y EL USO DE LA DISTRIBUCIÓN NORMAL

El propósito de la teoría del muestreo es que éste sea más eficiente. Su objetivo es desarrollar métodos de selección de muestras y de estimación, que propone, al menor costo posible, estimaciones con la suficiente exactitud específica a costo mínimo aparece una y otra vez en la presentación de la teoría. Para aplicar este principio debemos ser capaces de predecir en cualquier método de muestreo que se considere, la precisión y el costo esperados.

Sobre la precisión, no es posible predecir cuál será el error de una estimación en una situación específica porque esto implicaría el conocimiento del verdadero valor de la población. En lugar de ello, se juzga al examinar la distribución de frecuencia generada para las estimaciones, suponiendo que el proceso de muestreo se aplica varias veces a la misma población. Lo anterior es así porque cuando calculamos una muestra para algún caso de estudio con, por ejemplo, un 95% de confianza, lo que estamos diciendo es que, si se seleccionan 100 muestras diferentes, en el 95% de ellas vamos a obtener valores similares a los que obtuvimos con nuestra muestra, la muestra seleccionada por el investigador. Por eso, para efectos de calcular una muestra siempre es necesario tener información clave sobre las variables que deseamos estudiar, para realizar cálculos que se ajusten a la realidad, es decir, lo cual tiene como condición inexorable que la muestra seleccionada sea representativa de la población y, para ello, que los elementos de la muestra sean independientes entre sí o, al menos, intercambiables de alguna manera¹. En este caso, lo mínimo es poder ver cuál fue la varianza y

¹ "First, the straightforward summation of individual deviations, either $|y_1 - \hat{y}_i|$ or $(y_1 - \hat{y}_i)^2$, each depending on only one observation, implies that the observations are all made on the same physical scale and suggest that the observations are independent, or at least that they are in some sense exchangeable, so justifying an even-handed treatment of the components." (McCullagah & Nelder, 1989, pág. 5). Esto, como se señala en (Cochran, 1991, pág. 31), se traduce en el contexto de la teoría estándar de las encuestas por muestreo, como la *selección aleatoria* de los formularios (que es una aplicación del muestreo de probabilidad), la cual captura la noción de que cada unidad de población (cada formulario devuelto después del censo) tiene la misma oportunidad de ser incluida en la muestra.

promedio en estudios anteriores con la finalidad de ser considerados al construir nuestra muestra.

Una simplificación adicional que podemos hacer, consiste en suponer, lo que es razonable en la práctica si se trata de muestras de tamaño común², que las estimaciones de muestra siguen una distribución aproximadamente normal.

Existen dos diferencias entre la teoría estándar de encuestas por muestreo y la teoría clásica de muestro como aparece en los libros de estadística matemática. La primera diferencia es que, en el contexto de la teoría clásica, las mediciones hechas sobre las unidades de muestreo de la población suelen suponerse que siguen en conjunto una distribución de frecuencia de forma matemática conocida³, como sería la distribución normal, cuyos parámetros, media y varianza, por ejemplo, se estimarían a partir de los datos de las muestras. La segunda diferencia es que las poblaciones en una encuesta tienen un número finito de unidades. Los resultados son ligeramente más complicados cuando el muestreo es de una población finita y no de una infinita. Por razones prácticas, a menudo se ignoran estas diferencias en los resultados para poblaciones finitas e infinitas.

Considérese las siguientes condiciones. Un conjunto de muestras distintas se denota como S_1, S_2, \dots, S_v . Así, el investigador debe poder establecer con precisión cuáles son las unidades de muestreo que pertenecen a S_1, S_2, \dots, S_v . Cada muestra posible S_i tiene asignada una probabilidad de selección π_i y se selecciona una de las S_i por un proceso aleatorio, en el que cada S_i tiene una probabilidad π_i de ser elegida. Finalmente, el método para calcular la estimación a partir de la muestra debe ser definido y debe conducir a una estimación única para cualquier muestra

² Con *muestras de tamaño común* se refiere a los tamaños de muestra tradicionales para cada área de estudio en concreto, es decir, a los tamaños de muestra que en cada área de estudio se consideran (con base en la experiencia profesional) suficientes para la verificación del teorema central del límite en cada una de estas áreas de estudio. Por ejemplo, para una muestra pequeña un tamaño común es alrededor de 400 entrevistas.

³ Esto es, se conoce su distribución de probabilidad.

específica. Podemos decir, por ejemplo, que la estimación es el promedio de las mediciones correspondientes a las unidades individuales de la muestra.

Si un procedimiento de muestreo satisface la totalidad del conjunto de condiciones anteriormente definidas, es posible calcular la distribución de frecuencia de las estimaciones que genera el proceso, si se aplica repetidamente a la misma población. Sabemos la frecuencia con que se elige cualquier muestra S_i , y sabemos cómo calcular la estimación a partir de los datos de S_i . Por tanto, es claro que se puede desarrollar una teoría de muestreo para cada procedimiento de este tipo, aunque los detalles del desarrollo puedan ser intrincados. Un método de esta clase se conoce con el nombre de *muestreo probabilista*.

Es útil usar la palabra *estimador* para designar la regla por la cual se calcula alguna característica μ de la población, a partir de los resultados de la muestra⁴, y la palabra *estimación* para el valor obtenido de una muestra específica. Un estimador $\hat{\mu}$ de μ dado por un plan de muestreo se llama insesgado si el valor medio de $\hat{\mu}$, tomado sobre todas las muestras posibles por el plan, es igual a μ .

$$E(\hat{\mu}) = \sum_{i=1}^v \pi_i \hat{\mu}_i = \mu$$

En la expresión anterior, $\hat{\mu}_i$ es la estimación dada por la i -ésima muestra. Así, esto significa que el valor esperado, esperanza matemática o valor promedio de la característica μ que se desea estimar de la población (a partir de la muestra), es igual a la suma de todos los valores que toma la característica investigada (el estimador) para cada observación⁵ de la muestra (*i.e.*, $\hat{\mu}_i$), ponderándose cada una de ellas por su probabilidad de ocurrencia.

⁴ Resultados generados por la aplicación de algún modelo estadístico sobre la muestra previamente extraída de la población. De hecho, "la regla" por la cual se calcula alguna característica μ de la población es precisamente algún modelo estadístico.

⁵ Cada observación individualmente analizada (por supuesto, una observación no necesariamente está conformada de únicamente un elemento, eso dependerá de la investigación en concreto de que se trate.

Para el caso de una distribución normal, el intervalo de confianza del estimador se construye de la siguiente manera⁶:

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

El componente $\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ es el límite inferior del intervalo, mientras que $\bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ es el límite superior del intervalo. La magnitud $X\%$ ⁷ implica que, si se usara el mismo plan de muestreo muchas veces en una población, haciendo un enunciado de confianza⁸ a partir de cada muestra, aproximadamente el $X\%$ serían correctas y el $\left(1 - \frac{X\%}{100\%}\right) \times 100\%$ incorrectas⁹.

Con una variable normalmente distribuida, se usan tablas de distribución t de Student, en lugar de tablas normales para calcular los límites de confianza para μ cuando la muestra es pequeña. El reemplazo de la tabla normal por la tabla de t casi no hace diferencia si el número de grados de libertad en $\sigma_{\hat{\mu}}$ excede a 50. Con ciertos tipos de muestreo estratificado y con el método de muestreo repetido (véase la sección 11.19 de Cochran) los grados de libertad son pequeños y la tabla t es necesaria.

II. EL SESGO Y SUS EFECTOS

En algunos de los problemas más comunes, particularmente en la estimación de razones, se encuentra que los estimadores convenientes y apropiados son sesgados. Aun con los estimadores que son insesgados en muestreo probabilista, los errores de medición y las no-respuestas pueden producir sesgos en los números que se calculan a partir de los datos. Esto sucede, por ejemplo, cuando casi todas

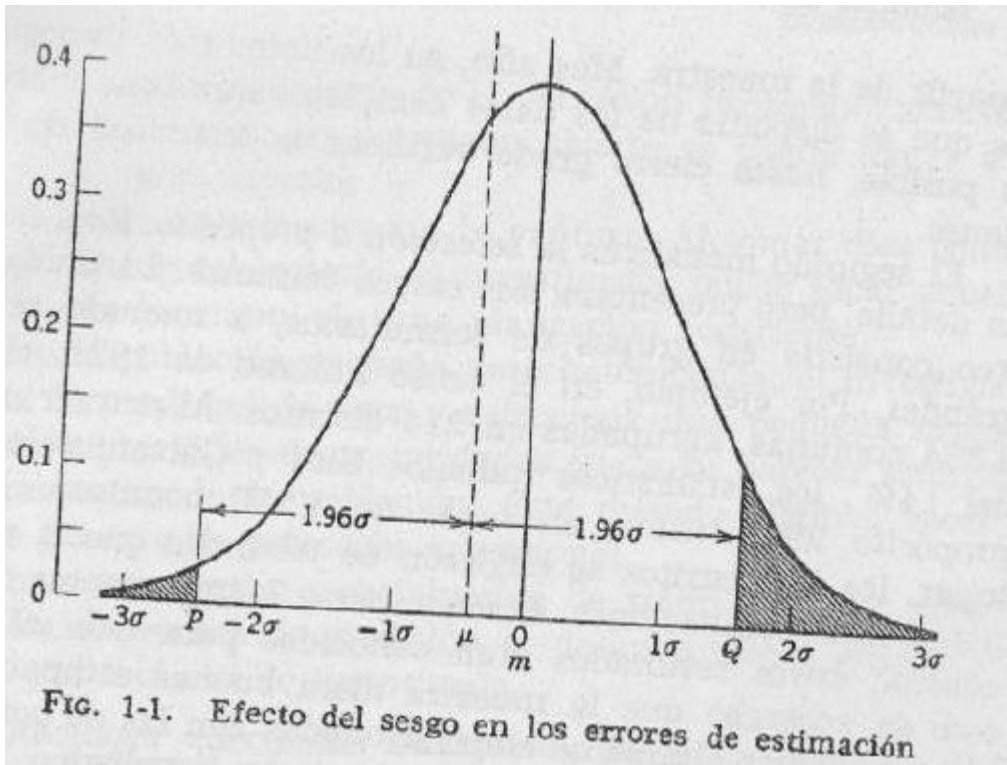
⁶ En donde $z_{\frac{\alpha}{2}}$ es el valor de Z en su tabla correspondiente, mientras que \bar{x} es la media muestral y n es el tamaño de la muestra.

⁷ Usualmente se trabaja con niveles de confianza del 90%, 95% y 99%, sin embargo, esto puede llegar a variar según la investigación.

⁸ Una afirmación que denote que una característica poblacional estimada (el estimador) se encuentra dentro de un rango de valores (el intervalo de confianza).

⁹ Aquí Cochran hace referencia a que el enunciado de confianza (véase la nota al pie anterior) se verificaría para el $X\%$ de los casos y no se verificaría para su complemento respecto del 100%.

las personas que se niegan a ser entrevistadas se oponen a cierto gasto de fondos públicos, en tanto que aquellas personas que han sido entrevistadas se dividen igualmente en pro y en contra. A continuación, se presenta una figura que resultará útil para estudiar el sesgo de un estimador.



Fuente: (Cochran, 1991, pág. 35).

La figura anterior, correspondiente a una distribución normal estandarizada, muestra que la estimación $\hat{\mu}$ está normalmente distribuida alrededor de una media m que está a una distancia B del verdadero valor μ de la población. La magnitud del sesgo o distancia B es $B = m - \mu$.

Supongamos, con la finalidad de comprender cómo el sesgo distorsiona el análisis probabilista, que se desconoce la existencia de dicho sesgo. Calculamos la desviación estándar σ de la distribución de frecuencias del estimador (esto será, desde luego, la desviación estándar respecto a la media m de la distribución, no

respecto a la verdadera media μ^{10} . Estamos usando σ en lugar de $\sigma_{\hat{\mu}}$. Como afirmación acerca de la exactitud de la estimación, declaramos que la probabilidad es de 0.05 (*i.e.*, α) de que el estimador $\hat{\mu}$ esté en error por más de 1.96σ .

Ahora es posible considerar cómo la presencia del sesgo distorsiona esta probabilidad. Para hacer esto calculamos la verdadera probabilidad de que la estimación está en error por más de 1.96σ , en donde el error es medido a partir de la verdadera media μ . Ambos lados de la distribución deben ser examinados separadamente. Para el extremo derecho, la probabilidad de un error de más de $+1.96\sigma$ es el área sombreada a partir de Q en la figura 1-1 expuesta anteriormente¹¹. Esta área se puede determinar cuantitativamente al resolver la siguiente integral:

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mu+1.96\sigma}^{\infty} e^{-\frac{(\hat{\mu}-m)^2}{2\sigma^2}} d\hat{\mu}$$

Sea $\hat{\mu} - m = \sigma t$, entonces despejando t se obtiene que $t = \frac{\hat{\mu}-m}{\sigma}$. Como $\hat{\mu}$ es la estimación de μ , entonces la relación teórica $\hat{\mu} = \mu$ es válida si y solo si no existe sesgo estadístico en el sentido definido por Cochran. Dando por sentado lo anterior, entonces $t = \frac{\hat{\mu}-m}{\sigma}$ es intercambiable a conveniencia por su versión teórica $t = \frac{\mu-m}{\sigma}$ y, por tanto, es posible escribir el límite de integración $\mu + 1.96\sigma$ como $\frac{\mu-m}{\sigma} + 1.96\sigma$. Si además se asume que la desviación estándar es unitaria, entonces $\frac{\mu-m}{\sigma} + 1.96\sigma$ se transforma en $\frac{\mu-m}{\sigma} + 1.96(1) = \frac{\mu-m}{\sigma} + 1.96$.

¹⁰ Nótese que Cochran denota la desviación muestral como σ en lugar del usual símbolo s , mientras que la desviación poblacional estimada la denota como $\sigma_{\hat{\mu}}$.

¹¹ En general, lo que en la figura mencionada se determina gráficamente es si la media m se encuentra dentro del intervalo de confianza de la media real μ , en donde este intervalo es dado por 1.96σ , lo cual es realizado a través del estimador $\hat{\mu}$, que es el estimador de la característica investigada de la población, cuyo valor real es μ .

Si se factoriza $\frac{\mu-m}{\sigma}$ como $\frac{-(-\mu+m)}{\sigma}$ y se recuerda que $B = m - \mu$, entonces el monomio anterior se puede expresar como $\frac{-B}{\sigma}$ puesto que $(-\mu + m) = [B = (m - \mu)]$. Así, $\frac{\mu-m}{\sigma} + 1.96$ puede expresarse como $\frac{-B}{\sigma} + 1.96$, lo que es equivalente a $1.96 - \frac{B}{\sigma}$.

Puesto que se estableció antes que $\sigma = 1$, se puede sustituir este valor en $\frac{1}{\sigma\sqrt{2\pi}}$, obteniéndose así $\frac{1}{\sqrt{2\pi}}$. Como $e^{-\frac{(\hat{\mu}-m)^2}{2\sigma^2}}$, $B = m - \mu$ y se ha asumido que no existe sesgo en la estimación, entonces $B = m - \mu$ es equivalente a $B = m - \hat{\mu}$ (puesto que $\hat{\mu} = \mu$ debido a la inexistencia de sesgo). Además, si se factoriza el exponente de $e^{-\frac{(\hat{\mu}-m)^2}{2\sigma^2}}$, puede verificarse que $-\frac{(\hat{\mu}-m)^2}{2\sigma^2} = -\frac{1}{2} \times \frac{(\hat{\mu}-m)^2}{\sigma^2}$ y también, como resultado de la propiedad de potencias $\frac{a^n}{b^n} = \left(\frac{a}{b}\right)^n$, que $-\frac{1}{2} \times \frac{(\hat{\mu}-m)^2}{\sigma^2} = -\frac{1}{2} \times \left(\frac{\hat{\mu}-m}{\sigma}\right)^2$. Como resultado de la inexistencia de sesgo, $t = \frac{\mu-m}{\sigma}$ puede escribirse como $t = \frac{\hat{\mu}-m}{\sigma}$ y, a su vez como resultado de ello, $-\frac{1}{2} \times \left(\frac{\hat{\mu}-m}{\sigma}\right)^2$ puede escribirse como $-\frac{1}{2} \times (t)^2$. Finalmente, $-\frac{1}{2} \times (t)^2$ puede escribirse como por lo que $e^{-\frac{(\hat{\mu}-m)^2}{2\sigma^2}}$ se transforma en $e^{-t^2/2}$, por lo que la integral anterior se puede volver a escribir, ahora en términos de t y bajo los supuestos ya definidos con antelación, como se presenta a continuación:

$$\frac{1}{\sqrt{2\pi}} \int_{1.96 - \frac{B}{\sigma}}^{\infty} e^{-t^2/2}, dt$$

De manera semejante, el extremo izquierdo de la figura 1-1, es decir, el área sombreada a partir de P , tiene un área dada por la integral:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1.96 - \frac{B}{\sigma}} e^{-t^2/2}, dt$$

Al observar las integrales, está claro que la perturbación depende únicamente de la razón del sesgo a la desviación estándar, es decir, de $\frac{B}{\sigma}$.

TABLA 1.1. EFECTO DE UN SESGO B EN LA PROBABILIDAD DE COMETER UN ERROR MAYOR QUE 1.96σ

B/σ	Probabilidad de error		Total
	$< -1.96\sigma$	$> 1.96\sigma$	
0.02	0.0238	0.0262	0.0500
0.04	0.0228	0.0274	0.0502
0.06	0.0217	0.0287	0.0504
0.08	0.0207	0.0301	0.0508
0.10	0.0197	0.0314	0.0511
0.20	0.0154	0.0392	0.0546
0.40	0.0091	0.0594	0.0685
0.60	0.0052	0.0869	0.0921
0.80	0.0029	0.1230	0.1259
1.00	0.0015	0.1685	0.1700
1.50	0.0003	0.3228	0.3231

Fuente: (Cochran, 1991, pág. 36).

En el caso de considerar la probabilidad total de cometer un error de más de 1.96σ , el sesgo tiene poco efecto, siempre y cuando sea menor que una décima de la desviación estándar, es decir, que el cociente $\frac{B}{\sigma}$ debe ser menor que 0.10. En el punto en el que $\frac{B}{\sigma} = 0.10$, el error es de 0.0511 en lugar de 0.05 (como se había definido antes al declarar la afirmación acerca de la exactitud de la estimación).

En la medida en que B (el sesgo) aumenta, la perturbación se vuelve más seria. Así, para $B = \sigma$, la probabilidad de error es de 0.17, más de 3 veces el valor supuesto (0.05). Los dos lados de la distribución se afectan de una manera diferente. Con un sesgo positivo, como en el ejemplo, la probabilidad de subestimación en más de 1.96σ disminuye con rapidez del valor supuesto de 0.025 (la distribución de 0.05

en dos colas, *i.e.*, $\frac{0.05}{2}$) para volverse despreciable cuando $B = \sigma$ (que implica que $\frac{B}{\sigma} = 1$), puesto que la probabilidad de error cuando $\frac{B}{\sigma} = 1$ es de 0.0015, mientras que lo contrario ocurre respecto a la probabilidad de sobreestimación en más de 1.96σ , que en el mismo escenario (cuando $\frac{B}{\sigma} = 1$) ya se vio al inicio de este párrafo que es de 0.17, y que en general se eleva de forma constante, como puede verse en la tabla 1.1. En la mayoría de aplicaciones el error total es de interés primordial, pero en ocasiones estamos particularmente interesados en los errores cometidos en una sola dirección.

Como una regla de trabajo, el efecto del sesgo en la exactitud de un estimador es despreciable si el sesgo es menor que un décimo de la desviación estándar de la estimación. Si tenemos un método de estimación que produce sesgo para el cual $\frac{B}{\sigma} < 0.1$, en donde B es el valor absoluto del sesgo, se puede argumentar que el sesgo no es una desventaja del método. Aún con $\frac{B}{\sigma} = 0.2$ es modesta la perturbación en la probabilidad de error.

En el uso de estos resultados se deben distinguir las dos fuentes de sesgo antes mencionadas, *i.e.*, los sesgos surgidos en la estimación de razones (proporciones) y los sesgos surgidos por errores de medición y no-respuestas. Con sesgos que surgen en la estimación de proporciones, un límite superior a la proporción $\frac{B}{\sigma}$ se puede determinar matemáticamente. Si la muestra es lo suficientemente grande, podemos tener la confianza de que $\frac{B}{\sigma}$ no excederá 0.1. Por otro lado, con sesgos causados por errores de medición y de no-respuesta, generalmente es imposible encontrar el límite superior para $\frac{B}{\sigma}$ que es pequeño. Sobre este enfadoso problema se discute en el capítulo 13 del libro “Técnicas de Muestreo” de Cochran.

III. ERROR CUADRÁTICO MEDIO

A fin de comparar un estimador sesgado, con un estimador insesgado, o bien, dos estimadores que tienen cantidades diferentes de sesgo, un criterio útil es el del error cuadrático medio (ECM) del estimador, medido a partir del valor de la población que se está estimando.

$$\text{ECM}(\hat{\mu}) = E(\hat{\mu} - \mu)^2 = E[(\hat{\mu} - \mu) + (m - \mu)]^2$$

$$\text{ECM}(\hat{\mu}) = E(\hat{\mu} - m)^2 + 2(m - \mu)E(\hat{\mu} - m) + (m - \mu)^2$$

$$\text{ECM}(\hat{\mu}) = [(varianza de \hat{\mu}) + (sesgo)^2]$$

Como señala (Cochran, 1991, pág. 37), el doble producto $2(m - \mu)E(\hat{\mu} - m)$ desaparece porque $E(\hat{\mu} - m) = 0$, es decir, porque se asume que la media muestral converge a la media poblacional o, lo que es lo mismo, que la esperanza matemática de su diferencia (o diferencia promedio) es nula.

El uso del ECM, como un criterio de la exactitud de un estimador, equivale a considerar dos estimadores que tienen el mismo ECM como equivalentes. Esto no es estrictamente correcto porque las distribuciones de frecuencia de los errores $(\hat{\mu} - \mu)$, de tamaños diferentes no serán las mismas para los dos estimadores si tienen cantidades diferentes de sesgo. Sin embargo, se ha demostrado por Hansen, Hurwitz y Madow (1953) que si $\frac{B}{\sigma}$ es aproximadamente menor que $\frac{1}{2}$, entonces las dos distribuciones de frecuencia son casi idénticas en relación con los errores absolutos $|\hat{\mu} - \mu|$ de magnitudes diferentes. La tabla 1.2 presentada a continuación ilustra este resultado.

TABLA 1.2. PROBABILIDAD DE UN ERROR ABSOLUTO $\geq 1\sqrt{ECM}$, $1.96\sqrt{ECM}$ y $2.576\sqrt{ECM}$

B/σ	Probabilidad		
	$1\sqrt{ECM}$	$1.96\sqrt{ECM}$	$2.576\sqrt{ECM}$
0	0.317	0.0500	0.0100
0.2	0.317	0.0499	0.0100
0.4	0.319	0.0495	0.0095
0.6	0.324	0.0479	0.0083

Fuente: (Cochran, 1991, pág. 37).

Aún si $\frac{B}{\sigma} = 0.6$, los cambios en las probabilidades al compararlas con los valores obtenidos para $\frac{B}{\sigma} = 0$ son ligeros.

Debido a la dificultad de poder asegurar que ningún sesgo insospechado se introduce en las estimaciones, generalmente hablaremos de *precisión* de un estimador más que de su exactitud. La exactitud se refiere a la magnitud de las desviaciones respecto a la media verdadera μ , mientras que precisión se refiere a la magnitud de las desviaciones respecto a la media m , obtenida por la aplicación repetida del procedimiento de muestreo.

IV. MUESTREO ALEATORIO SIMPLE

IV.I. Generalidades

Debe comenzarse diciendo que el muestreo aleatorio simple es un tipo de muestreo de probabilidad. Formalmente, es un método de selección de n unidades en un conjunto de N de tal modo que cada una de las C_N^n muestras tengan la misma oportunidad de ser seleccionadas, en donde $C_N^n = \frac{n!(N-n)!}{N!}$. La expresión C_N^n denota las posibles muestras de tamaño n de una población N , *i.e.*, las distintas combinaciones de n elementos que es posible hacer de un determinado conjunto conformado por N elementos.

Por supuesto, deben declararse los supuestos detrás de la construcción de muestras antes descrita, los cuales son que cada muestra se debe diferenciar en al menos uno de sus elementos, que el orden no importa y que los elementos de una muestra no se pueden repetir en otra muestra.

En la práctica, un muestreo aleatorio se realiza unidad por unidad. Se numeran las unidades entre 1 y N . Posteriormente se extrae una serie de n números aleatorios entre 1 y N , ya sea utilizando una tabla de números aleatorios o mediante un programa de computación que produce una tabla semejante. En cada extracción, el proceso debe otorgar la misma oportunidad de selección a todos y cada uno de los números que *no hayan salido*. Las unidades que llevan estos n números constituyen la muestra.

Fácilmente se verifica que todas las C_N^n muestras distintas, tienen la misma oportunidad de ser extraídas por este método. Considérese una muestra determinada, es decir, una colección de n unidades especificadas. En la primera extracción, la probabilidad de que se seleccione una de estas n unidades es de $\frac{n}{N}$. En la segunda, la probabilidad que se extraiga una de las restantes $(n - 1)$ unidades especificadas es $\frac{n-1}{N-1}$, y así sucesivamente. Por lo tanto, la probabilidad de que se extraigan las n unidades especificadas es:

$$\frac{n}{N} \times \frac{n-1}{N-1} \times \frac{n-2}{N-2} \cdots \frac{1}{N-n+1} = \frac{N!}{n!(N-n)!} = \frac{1}{C_N^n}$$

Como en todas las extracciones subsecuentes se descarta un número extraído, este método también se llama *muestreo aleatorio sin restitución* o *muestreo aleatorio sin reemplazo*. El muestreo aleatorio *con restitución* es perfectamente factible: en cada extracción todos los N miembros de la población reciben la misma oportunidad de extracción, sin que importe el número de veces que se extrajeron antes. Las fórmulas de varianzas y varianzas estimadas de las estimaciones realizadas a partir de la muestra son a menudo más simples cuando el muestreo es con restitución

que en el caso contrario. Por esta razón, se utiliza el muestreo con restitución en los planes de muestreo más complicados, aunque a primera vista parece inútil tener dos o más veces la misma unidad dentro de la muestra.

IV.II. Selección de una Muestra Aleatoria Simple

Las tablas de números aleatorios son tablas de dígitos 0, 1, 2, ..., 9, donde cada dígito tiene la misma probabilidad de ser seleccionado en cada extracción. Cuando se usan tablas para seleccionar una muestra aleatoria simple, el primer paso es la enumeración de las unidades de la población de 1 a N . Si el primer dígito de N es un número entre 5 y 9, el método de selección presentado a continuación es el adecuado.

TABLA 2.1. UN MILLAR DE DÍGITOS ALEATORIOS

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067

Fuente: (Cochran, 1991, pág. 43).

Supóngase $N = 528$ y queremos $n = 10$. Tómense tres columnas de la tabla 2.1, digamos la 25, 26 y 27. Recórrase hacia abajo cada columna seleccionando los 10 primeros números *distintos*, entre 001 y 528. Tales números son 36, 509, 364, 417, 348, 127, 149, 186, 290 y 162. Para los dos últimos números saltamos a las columnas 30, 31 y 32, puesto los números entre 001 y 528 se agotan en las columnas inicialmente seleccionadas. En selecciones repetidas es aconsejable cambiar el punto de partida de la tabla.

A continuación, se señala el bloque correspondiente a las columnas 25, 26 y 27 en un rectángulo rojo de líneas gruesas, los números seleccionados en rectángulos rojos de líneas delgadas y los números que no se seleccionaron por exceder 528 en rectángulos anaranjados.

TABLA 2.1. UN MILLAR DE DÍGITOS ALEATORIOS

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67332	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16259	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84757	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76756	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067

Fuente: (Cochran, 1991, pág. 43).

La desventaja de este método es que los números de tres dígitos 000 y todos los del 529 al 999 no se usan, aunque al saltarse números no se pierde mucho tiempo. Cuando el primer dígito de N es menor que 5, algunos pueden preferir este método si n es pequeño y se dispone de una tabla con cantidad de números aleatorios bastante grande. Con $N = 128$, por ejemplo, un segundo método con menos rechazos y de más fácil aplicación es el siguiente. En una serie de números de tres dígitos se subtrae 200 de todos los números que hay entre 201 y 400, se subtrae 400 de todos los números entre 401 y 600, 600 de todos los números entre 601 y 800, 800 de todos los números entre 801 y 999 y desde luego 000 de todos los números entre 000 y 200. Todos los residuos mayores que 129 y los números 000, 200, etc., se desechan. Al utilizar las columnas 05, 06 y 07 de la tabla 2.1 obtenemos 26, 52, 7, 94, 16, 48, 41, 80, 128 y 92, la extracción que requiere de 15 números de tres dígitos para $n = 10$.