Revista Colombiana de Estadística
Volumen 24 (2001) N° 1, páginas 27 *a* 32

UN TAMAÑO DE MUESTRA PRELIMINAR EN LA ESTIMACION DE LA MEDIA, EN POBLACIONES CON DISTRIBUCIONES UNIFORMES Y TRIANGULARES

CARLOS E. ALONSO M. *
DAVID OSPINA B.**

Resumen

En la adecuada selección de una muestra aleatoria simple, uno de los aspectos contemplados se refiere al tamaño mínimo requerido para garantizar unos niveles de confiabilidad y precisión, establecidos de antemano, en las estimaciones. En el proceso de estimación de la media poblacional, con la mesia muestral como estimador, es necesario el conocimiento previo del valor de la varianza, σ^2. Si este no se tiene, se toma una muestra preliminar cuyo objetivo es obtener una estimación de σ^2. En este trabajo se presenta una metodología adecuada para estimar este tamaño de muestra preliminar para el caso de poblaciones cuyas distribuciones pueden ser aproximadas a una distribución uniforme o triangular. La metodología puede aplicarse a muchas otras situaciones.

Palabras Claves: Muestra Aleatoria Simple, Muestra Preliminar, Distribuciones Uniforme, Triangular y Gamma, Simulación de Monte Carlo.

*Instructor Asociado, Departamento de Estadística, Universidad Nacional de Colombia
**Profesor Asociado, Departamento de Estadística, Universidad Nacional de Colombia

1. Introducción

En el proceso de selección de una muestra aleatoria simple se contemplan por lo menos dos aspectos de importancia a saber: el procedimiento de escogencia de los individuos o unidades a observar, y el tamaño de muestra mínimo para garantizar un nivel de confianza y una precisión fijados de antemano [Clavijo et al, 1995]. Este artículo trata el segundo de estos aspectos.

El cálculo del tamaño de muestra final para estimar la media poblacional implica conocer el valor de la varianza poblacional σ^2. Si este valor no se conoce o su conocimiento es muy vago, "como sucede en la mayoría de casos [Clavijo et al,1995]", se recomienda usar la cuasivarianza muestral calculada a partir de una muestra preliminar.

La preocupación por el tamaño de la muestra preliminar no es nueva. Ya en 1945 Stein presenta un procedimiento de dos etapas para estimación de la media en poblaciones normales con varianza desconocida [Govindarajulu, 1975], donde demuestra que el tamaño de la segunda muestra depende de los resultados de la primera. Este procedimiento a dos etapas es un caso particular del Análisis Secuencial desarrollado por Wald (1945). Cochran [1977] muestra la importancia del tamaño de la muestra final cuando afirma que "para cada plan (de selección de muestra), que sea considerado, rigurosas estimaciones del tamaño de muestra pueden ser hechas a partir del conocimiento del grado de precisión deseado". El error tolerado depende directamente del valor de la varianza, valor que es estimado a partir de una muestra preliminar.

En este trabajo se parte de un tamaño de muestra preliminar igual a 10, sugerido por la experiencia, para la estimación de σ^2, y se presenta una sustentación para una elección del tamaño de muestra preliminar en poblaciones cuya distribución se pueda aproximar a una distribución uniforme o a una distribución triangular, basándose en resultados obtenidos mediante simulación.

Documentos que presentan información de relevancia con el tema tratado son: el articulo de Lusk y Wright(1982) donde se presenta un procedimiento para derivar la función de densidad de probabilidad para la suma de variables aleatorias con distribución uniforme utilizando la transformación inversa de Laplace, método no trivial que se complica cuando el número de variables crece; y el artículo de Clavijo, J. et al. (1995), donde se plantean tamaños de muestra mínimos para lograr una aproximación adecuada a la distribución normal para poblaciones simétricas acotadas unimodales.

2. Metodología de Trabajo

El problema central consiste en determinar el tamaño de muestra n para estimar la media poblacional μ, con un nivel de confianza $1-\alpha$, $0<\alpha<1$, y un error absoluto de muestreo ϵ, con $\epsilon>0$, determinado. Si σ^2 es conocido, n puede calcularse con fórmulas bien definidas. Para el caso en el cual σ^2 es desconocido, es conveniente seleccionar una muestra preliminar de tamaño m, para obtener a partir de ella una estimación de σ^2. Obtenida ésta, se calcula el valor de n y luego se selecciona el número restante de observaciones, $n-m$, necesarias para lograr la estimación deseada.

Dado que el interés de la muestra preliminar se centra en la estimación de σ^2 la mayor parte del trabajo gira en torno a la distribución de su estimador S^2 (varianza muestral). El trabajo realizado para cada una de las distribuciones, Uniforme y Triangular, es presentado en el siguiente algoritmo:

1. Simulación de k muestras de tamaño m de la distribución correspondiente, con parámetros 0 y 1. En este trabajo el valor de k es 1000 y el tamaño preliminar m inicial es 10.

2. Cálculo de las estimaciones de la varianza poblacional σ^2 y estimación de la distribución del estimador S^2. En ese caso la mejor distribución ajustada encontrada es una distribución gamma cuyos parámetros se estiman usando el primer y segundo momentos muestrales.

3. Establecimiento de un intervalo de confianza unilateral (inferior) del 95 % para σ^2, cuyo límite es el percentil cinco de la distribución gamma ajustada.

4. Cálculo del tamaño de muestra final n, tomando como base el percentil cinco mencionado en el tercer item [1].

5. Simulación de l muestras de tamaño n. En este caso l se fijó en 1000.

6. Estimación de la media poblacional para cada una de las muestras simuladas mediante un intervalo de dos colas. Si el 95 % o más de estos intervalos cubre la media verdadera, se acepta que m es un tamaño adecuado para la muestra preliminar. En caso contrario se vuelve al primer punto, haciendo $m=m+1$. El proceso se repite iterativamente cuantas veces sea necesario.

[1] Para el cálculo de n se utiliza la ecuación $n=\left(z^2_{1-\frac{\alpha}{2}}(\frac{S}{\bar{y}})^2\right)/\epsilon_r^2$ donde $\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$ y $s^2=(\sum_{i=1}^{n}(y_i-\bar{y})^2)/(n-1)$, $Z_{1-\frac{\alpha}{2}}$ percentil $100(1-\frac{\alpha}{2})$ de la distribución normal estándar, ϵ_1 error relativo o error absoluto de muestreo dividido por $\bar{y}$

7. Generalización para distribuciones *Uniforme*$(0, b)$ y *Triangular*$(0, b)$. Los valores de b usados son $1, 2, 3, 4, 5, 10, 20, 50, \cdots, 1{,}000{,}000$

3. Resultados

De acuerdo a las tablas que se presentan al final del artículo, Tabla No 1.U(0,b) y Tabla No 2. T(0,b), se puede concluir, para todos los valores del parámetro b, que al menos 95 % de los intervalos de confianza calculados cubren el verdadero parámetro poblacional, utilizando un tamaño de muestra preliminar igual a 10.

4. Conclusiones

De acuerdo a la metodología usada un tamaño de muestra preliminar igual a 10 es adecuado como tamaño de muestra preliminar en la estimación de la media en poblaciones con distribuciones cercanas a una distribución uniforme o triangular. La metodología utilizada en este trabajo puede ser aplicada en forma amplia a otras distribuciones bajo condiciones regulares, como son la existencia de primeros momentos con respecto a la media.

Referencias

[1] Clavijo, J. A. - Ospina (1995). *D. Normal aproximation when sampling from bounded symmetric distributions*, Revista Colombiana de Estadística, 31 y 32, 23-30, Bogotá, Colombia.

[2] Cochran, W. (1977). *Sampling Techniques.* John Wiley & Sons Inc., New York, 3a edición.

[3] Govindarajulu Z (1975). *Sequential statistics procedures.* Academic Press, New York, San Francisco.

[4] Lusk, E. J. - Wright, H. (1982). *Deriving the probability density for sums of uniform random variables*, JASA, Mayo Vol 36, No 2

Resultados de Distribución Uniforme									
Parámetro				Parámetros Gamma		Percentil 5	n	Intervalos que	Media
a	b	m	$p - valor^2$	α	β	$G(\alpha, \beta, 0{,}05)$	Final	Incluyen la Media	Poblacional
0	1	10	0.419	9.74	115.410	4.527E-02	279	960	0.5
		12	0.467	12.16	144.000	4.776E-02	294	956	
0	2	10	0.281	8.73	25.783	1.745E-01	269	958	1
		11	0.668	10.55	31.427	1.857E-01	286	955	
0	3	10	0.463	9.50	12.455	4.061E-01	278	962	1.5
		11	0.422	10.73	14.021	4.250E-01	291	956	
0	4	10	0.419	9.34	6.970	7.094E-01	273	959	2
		11	0.517	10.65	7.934	7.446E-01	287	954	
0	5	10	0.537	9.13	4.370	1.096E+00	270	962	2.5
		11	0.453	9.66	4.570	1.133E+00	279	960	
0	10	10	0.433	9.52	1.129	4.493E+00	277	955	5
		11	0.625	11.44	1.363	4.767E+00	293	960	
0	20	10	0.302	10.27	0.310	1.815E+01	279	960	10
		11	0.554	9.90	0.292	1.831E+01	282	963	
0	50	10	0.449	8.62	0.041	1.068E+02	263	960	25
		11	0.732	10.48	0.049	1.174E+02	289	956	
0	100	10	0.493	9.22	0.001	4.405E+02	271	956	50
		11	0.512	10.41	0.013	4.529E+02	279	960	
0	200	10	0.281	9.23	0.003	1.793E+03	276	965	100
		11	0.344	9.46	0.003	1.785E+03	275	959	
0	1000	10	0.786	9.10	0.0001	4.384E+04	270	964	500
		11	0.842	10.74	0.13E-3	4.712E+04	290	962	
0	5000	10	0.544	8.76	4.15E-06	1.091E+06	269	958	2500
		11	0.866	10.03	4.70E-06	1.149E+06	283	953	
0	10000	10	0.580	9.45	1.12E-06	4.477E+06	276	965	5000
		11	0.674	9.54	1.13E-06	4.502E+06	277	955	
0	50000	10	0.447	8.81	4.11E-08	1.108E+08	273	959	25000
		11	0.443	9.66	4.53E-08	1.143E+08	281	963	
0	100000	10	0.395	8.52	1.00E-08	4.331E+08	267	959	50000
		11	0.635	9.40	1.10E-08	4.518E+08	278	962	
0	500000	10	0.318	9.40	4.50E-10	1.108E+10	273	959	250000
		11	0.365	10.73	5.00E-10	1.193E+10	294	956	
0	1000000	10	0.380	8.72	1.04E-10	4.324E+10	266	957	500000
		11	0.416	11.10	1.35E-10	4.618E+10	284	964	

Resultados de Distribución Triangular									
Parámetro				Parámetros Gamma		Percentil 5	n	Intervalos que	Media
a	b	m	$p - valor^2$	α	β	$G(\alpha,\beta,0{,}05)$	Final	Incluyen la Media	Poblacional
0	1	10	0.176	5.81	137.01	0.01809	112	955	0.5
		11	0.643	6.41	154.96	0.01832	113	951	
0	2	10	0.583	5.75	3.7	0.07247	112	950	1
		11	0.653	6.90	41.76	0.0769	119	951	
0	3	10	0.950	5.90	15.31	0.1661	114	961	1.5
		11	1.000	6.79	17.96	0.1743	120	962	
0	4	10	0.831	5.98	8.845	0.2938	113	950	2
		11	0.721	6.29	9.421	0.2978	115	957	
0	5	10	0.935	5.63	5.405	0.4387	108	954	2.5
		11	0.750	6.50	6.24	0.4714	116	951	
0	10	10	0.858	5.70	1.375	1.7547	108	971	5
		11	0.687	6.28	1.51	1.8534	114	950	
0	20	10	0.998	5.58	0.335	6.972	108	952	10
		11	0.906	6.95	0.414	7.854	121	953	
0	50	10	0.601	5.80	0.055	45.443	112	953	25
		11	0.377	6.00	0.0559	46.755	115	954	
0	100	10	0.858	5.93	0.014	179.349	111	958	50
		11	0.889	6.27	0.015	185.350	114	954	
0	200	10	0.704	6.04	3.55E-03	743.691	115	952	100
		11	0.864	6.19	3.77E-03	726.890	112	955	
0	500	10	0.865	5.94	5.75E-04	4471.670	110	951	250
		11	0.843	6.46	6.15E-04	4745.320	117	958	
0	1000	10	0.886	5.91	1.42E-04	1.80E+04	111	954	500
		11	0.863	6.75	1.59E-04	1.96E+04	121	950	
0	5000	10	0.984	5.92	5.51E-06	4.65E+05	115	954	2500
		11	0.981	6.68	6.38E-06	4.82E+05	119	950	
0	10000	10	0.747	5.60	1.36E-06	1.73E+06	107	953	5000
		11	0.643	6.85	1.65E-06	1.92E+06	119	951	
0	20000	10	0.906	5.69	3.40E-07	7.10E+06	110	955	10000
		11	0.844	6.60	3.94E-07	7.66E+06	118	951	
0	50000	10	0.581	5.77	5.52E-08	4.47E+07	110	952	25000
		11	0.711	6.88	6.44E-08	4.97E+07	123	957	
0	100000	10	0.816	5.40	1.13E-08	1.73E+08	107	955	50000
		11	0.858	6.72	1.60E-08	1.93E+08	119	952	
0	500000	10	0.959	6.22	5.98E-10	4.61E+09	114	954	250000
		11	0.983	6.83	6.42E-10	4.93E+09	122	952	
0	1000000	10	0.903	6.29	1.52E-10	1.84E+10	113	954	500000
		11	0.751	6.35	1.52E-10	1.87E+10	115	950	