

# Un análisis estadístico de muestras combinadas

Roberto A. Molina Cruz

Para monitorear la fortificación de la sal –entre otras cosas– se estableció en 1994 el Programa de Escuelas Centinelas de Micronutrientes, como un esfuerzo conjunto del Ministerio de Educación, El Fondo de las Naciones Unidas para la Infancia (UNICEF) y el Instituto de Nutrición de Centroamérica y Panamá (INCAP). Desde entonces este programa ha venido encuestando cada año a alumnos de escuelas primarias rurales del país y analizando muestras de sal de sus hogares. Uno de los principales objetivos de este programa es estimar la cantidad de yodo que los habitantes del país –especialmente los niños– consumen con la sal.

El Programa visita alrededor de 400 escuelas rurales de todo el país. En cada una de las cuales encuesta a 20 niños, y les pide una muestra de aproximadamente 5 gramos de sal de sus hogares. Como el análisis químico de estas muestras es costoso y requiere de por lo menos 50 gramos del alimento para medir la cantidad de yodo que contiene,

con las 20 muestras individuales se obtienen 2 muestras combinadas (en inglés: *pooled samples*) conteniendo 10 muestras individuales cada una. En el laboratorio se determina el número de partes del micronutriente por millón de partes del alimento (ppm) de cada muestra combinada, y hasta hace algunos años el Programa reportaba estos resultados a nivel nacional y de cada departamento del país.

En el año 2000 el jefe de la Unidad de Análisis y Protección de Alimentos del INCAP solicitó al Departamento de Matemática de la Universidad del Valle de Guatemala (UVG) analizar la encuesta. Durante ese mismo año se presentó un reporte que entre otras cosas señala las dificultades del análisis estadístico de las muestras combinadas, las cuales son relativamente comunes en el estudio de los alimentos y sus micronutrientes, así como del medio ambiente y sus contaminantes. Desde luego, en el reporte sugerimos hacer modificaciones a algunos de los aspectos de la encuesta.

En un artículo anterior (Molina et al, 2001) se discutieron los aspectos propios del procedimiento de muestreo utilizado. Especialmente, proponemos un procedimiento adecuado, sugerimos un método composición de las muestras considerando a los niños encuestados con hermanos en la misma escuela, el tamaño de las familias de los niños, y mostramos la necesidad de aplicar metodología de encuestas con varios marcos de muestreo.

En este artículo advierto sobre el cuidado que debe tenerse en el análisis estadístico de las muestras combinadas, en particular para estimar la distribución de las personas de una población respecto la cantidad de yodo que consumen con la sal. Aunque aquí menciono solamente el caso de la sal y el yodo, esta discusión es válida para cualquier alimento y micronutriente. Específicamente, (a) presento un modelo de la población en estudio, (b) explico por qué es necesario un análisis adecuado de los resultados de las muestras combinadas, y (c) discuto dos métodos para estimar los parámetros del modelo presentado.

### Un modelo de la población en estudio

Si consideramos la población en estudio formada por todas las comunidades de las escuelas y suponemos que cada comunidad puede tener básicamente dos fuentes de sal: los productores que fortifican la sal y los que no lo hacen, entonces podemos ver cada comunidad dividida en dos grupos de personas: las personas que consumen sal fortificada y las que consumen sal sin fortificar. Este primer grupo de personas pueden constituir la fracción:  $1-\lambda_k$ , digamos de

la  $k$ -ésima comunidad, y desde luego el segundo grupo la fracción:  $\lambda_k$ .

Es muy probable que la sal fortificada de la comunidad sea producida por un solo productor, quien tiene establecido su proceso de fortificación para producir sal con una cantidad promedio de yodo:  $\mu_k$ , y con alguna varianza:  $\sigma_k^2$ . Por lo que podemos suponer que las personas de este primer grupo de la comunidad están distribuidas en forma Normal  $(\mu_k, \sigma_k^2)$ , con respecto a la cantidad de yodo que consumen con la sal. Y que todas las personas de la comunidad están distribuidas en forma normal truncada con una masa de probabilidad igual a  $\lambda_k$  en cero, y con las siguientes media y varianza –esto es haciendo uso de las probabilidades condicionadas.

$$v_k = (1 - \lambda_k) \mu_k$$

$$\tau_k^2 = (1 - \lambda_k) [\lambda_k \mu_k^2 + \sigma_k^2]$$

Ahora bien, es posible considerar un solo modelo de este tipo para las comunidades que forman un dominio de estudio –como podría ser un departamento, una región o todo el país. Para ver esto debemos notar que la ley requiere a todos los productores fortificar la sal con valores entre 30 y 100 ppm de yodo (Congreso de la República de Guatemala, 1995). Entonces es posible que los productores que proveen de sal fortificada a las personas de un dominio de estudio específico lo hacen con un mismo valor objetivo:  $\mu$ , el cual seguramente es un poco mayor a 30 ppm, tratando que por ejemplo 90% de su producción satisfaga este requerimiento mínimo. Desde luego, este porcentaje depende de la varianza del procedimiento de fortificación que utilicen, pero es lógico pensar que los productores emplean el procedimiento

con la menor varianza posible y que se ajusta a sus costos. Por lo que también es posible que (aunque con procedimientos diferentes) la sal está siendo fortificada con aproximadamente la misma varianza:  $\sigma^2$ .

Considerando estos supuestos, las comunidades de un mismo dominio de estudio podrían tener asociadas una misma media y varianza, digamos

$$\mu_k = \mu \quad \text{y} \quad \sigma_k^2 = \sigma^2$$

Y estas comunidades deberían solo diferir significativamente respecto a sus parámetros:  $\lambda_k$ , los cuales desde luego componen el parámetro del dominio de estudio:  $\lambda$ , como sigue.

$$\lambda = (1/N) \sum_{k=1, \dots, m} N_k \lambda_k$$

Donde  $N_k$  representa el número de habitantes de la comunidad  $k$ ,  $m$  el número de comunidades del dominio de estudio, y  $N = \sum_{k=1, \dots, m} N_k$ , el número total de habitantes del mismo dominio.

Luego bajo estos mismos supuestos, las personas del dominio de estudio podrían también estar distribuidas en forma normal truncada, con una masa de probabilidad igual a  $\lambda$  en cero, y con el siguiente modelo describiendo su media y varianza poblacionales.

$$\begin{aligned} v &= (1 - \lambda) \mu \\ \tau^2 &= (1 - \lambda) [\lambda \mu^2 + \sigma^2] \end{aligned}$$

## La necesidad de un análisis estadístico adecuado

Al mezclar las 20 muestras individuales de sal de cada escuela en dos muestras combinadas –como ha estado haciendo el Programa– el promedio del contenido de yodo de estas muestras combinadas:  $\gamma_k^{(1)}$  y  $\gamma_k^{(2)}$ , estima al parámetro  $v_k$  de la comunidad, mientras que su varianza puede usarse para estimar el parámetro  $\tau_k^2$ . Esto es,

$$v_{k,est} = (\gamma_k^{(1)} + \gamma_k^{(2)}) / 2$$

$$\tau_{k,est}^2 = 5 (\gamma_k^{(1)} - \gamma_k^{(2)})^2$$

Si suponemos por ejemplo que los productores que fortifican la sal tienen un valor objetivo:  $\mu_k = 45$  ppm, entonces empleando la primera ecuación del modelo de cada comunidad podemos estimar el porcentaje de las persona que están consumiendo sal fortificada. Esto es

$$1 - \lambda_{k,est} = v_{k,est} / 45.$$

En las primeras columnas de la Tabla 1 presento la distribución de las 276 escuelas encuestadas en 1999, con respecto al contenido promedio de yodo de sus muestras combinadas:  $v_{k,est}$ . Notemos especialmente que 142 (51.4%) comunidades presentaron muestras combinadas con promedios menores a 15 ppm. De hecho, las 276 (100%) muestras combinadas contenían algo de yodo, siendo la menor cantidad observada 0.3 ppm. Además que 113 (41.0%) comunidades presentaron muestras combinadas con promedios entre 15 y 30 ppm, y que las restantes 21 (7.6%) comunidades presentaron muestras combinadas con promedios

**Tabla 1**

Escuelas encuestadas en 1999.

$v_k, est$	Frecuencia	Porcentaje	$1-\lambda_k, est$ (con $\mu_k=45$ ppm)
0 – 15 ppm	142	51.4	0.00 – 0.33
15 – 30	113	41.0	0.33 – 0.66
30 – 45	19	6.9	0.66 – 1.00
45 – más	2	0.7	1.00
	276	100.0	

mayores a 30 ppm, siendo el mayor contenido de yodo observado 57.9 ppm.

Estos resultados no indican qué porcentaje de las personas de las comunidades están consumiendo sal sin fortificar, incluso pueden hacernos pensar que toda la sal está fortificada pero mucha o alguna con muy poco yodo. Sin embargo, con el modelo recién descrito podemos estimar dicho porcentaje. Aunque debemos notar que estas estimaciones corresponden al porcentaje de muestras individuales sin fortificar, pero asumiendo que los hogares de los niños tienen un número similar de miembros, estos mismos porcentajes corresponden al de las personas que consumen sal sin fortificar.

En la última columna de la misma Tabla 1, aparecen los correspondientes intervalos del porcentaje de las personas que consumen sal fortificada:  $1-\lambda_{k, est}$ . Notemos ahora que el porcentaje de personas que consumen sal con algo de yodo podría ser de entre 0% y 33% en las 142 (51.4%) comunidades antes mencionadas, de entre 33% a 66% en las siguientes 113 (41.0%) comunidades,

y de 66% o más en solamente las restantes 22 (7.6%) comunidades encuestadas.

Con el modelo descrito podemos además estimar la distribución de las personas de cada comunidad y cualquier dominio de estudio, con respecto a la cantidad de yodo que consumen con la sal. Considerando los mismos datos de 1999, en la Tabla 2 presento la estimación de la distribución de dos comunidades: A y B, junto con la estimación de la distribución de un dominio de estudio cuya muestra comprendió a 83 comunidades –escuelas. Notemos que en la comunidad A estimamos que 51.83% de las personas consume sal no fortificada, 23.77% de las personas consumen sal con yodo pero con hasta 30 ppm, y que solamente 24.40% de las personas están consumiendo sal con más de 30 ppm. Por el otro lado, en la comunidad B estimamos que solo 12.50% de las personas consumen sal no fortificada, 32.21% de las personas consumen sal con hasta 30 ppm de yodo, 55.29% de las personas

**Tabla 2**

Distribuciones por cantidad del yodo en la sal

	0	(0,15]	(15,30]	(30,60]	(60,∞)
Comunidad A:	51.83	5.84	17.93	23.77	0.63
Comunidad B:	12.50	14.80	17.41	32.21	23.08
Dominio:	49.71	6.69	13.54	20.23	9.83
(83 comunidades de la muestra)					

**Tabla 3**

Distribuciones de las muestras combinadas.

	0	(0,15]	(15,30]	(30,60]	(60,∞)
Muestras:	0.00	55.97	35.43	8.60	0.00

consumen sal con más de 30 ppm, y que incluso 23.08% de las personas consumen sal con más de 60 ppm de yodo.

En el dominio de estudio estimamos que 49.71% de las personas consumen sal no fortificada, 20.23% de las personas consumen sal fortificada pero con hasta 30 ppm de yodo, 30.06% de las personas consumen sal con más de 30 ppm, y que incluso 9.83% de las personas consumen sal con más de 60 ppm. Esto contrasta con la distribución de las 83 muestras combinadas analizadas que aparece en uno de los reportes del Programa (UNICEF, 1999) y que presentamos en la Tabla 3, en la cual observamos que 100% de las muestras contienen algo de yodo, que 55.97% de las muestras combinadas contienen hasta 15 ppm, 35.43% de

las muestras combinadas contienen entre 15 y 30 ppm, que 8.60% de las muestras combinadas contienen entre 30 y 60 ppm de yodo, y que ninguna muestra combinada contiene más de 60 ppm de yodo.

### La estimación de los parámetros del modelo

Para estimar la distribución de los habitantes de cada dominio de estudio con respecto a la cantidad de yodo que consumen con la sal, se deben estimar los parámetros:  $\lambda$ ,  $\mu$ , y  $\sigma$ , al nivel del país y de cada departamento. Y dado que la muestra de la encuesta es obtenida mediante un procedimiento complejo de muestreo –esto por que considera estratos, conglomerados y diferentes probabilidades de selección–

los estimadores que se deben emplear también son complejos. Sin embargo, una descripción detallada de estos estimadores va más allá del objetivo de este artículo, y de hecho el lector interesado puede referirse al artículo citado antes (Molina et al, 2001).

Resulta que estos estimadores complejos se basan en las estimaciones de los mismos parámetros pero al nivel de cada conglomerado considerado en la encuesta, estos son las comunidades de las escuelas encuestadas. Es decir, se basan en los parámetros antes mencionados:  $\lambda_k$ ,  $\mu_k$ , y  $\sigma_k$ , de cada una de las comunidades, en particular de la  $k$ -ésima comunidad. Por lo que en esta sección describimos 2 métodos que podían ser empleados para estimar estos parámetros.

Debemos notar que los cálculos descritos pueden usarse directamente para estimar la distribución de los habitantes con respecto a la cantidad de yodo que consumen con la sal, esto es si asumimos que las familias de los alumnos encuestados tienen un número similar de miembros. De no sumir esto entonces es necesario emplear fórmulas de cálculo un poco más complicadas, pero las cuales se basan en las mismas ideas aquí discutidas.

#### • Método 1

Existe una prueba rápida y poco costosa que permite evaluar si una muestra de sal contiene algo yodo. La cual al parecer tiene el único inconveniente de su limitada sensibilidad, en particular de no poder identificar muestras de sal conteniendo cantidades muy pequeñas de yodo. Sin embargo, esta misma prueba –o alguna prueba mejor– podría

aplicarse a las muestras individuales de cada escuela, esto para separar las muestras que contienen algo de yodo de las muestras sin yodo, y luego componer las muestras combinadas con solamente las muestras individuales con yodo.

Entonces, los parámetros de la  $k$ -ésima comunidad pueden estimarse como describimos a continuación.

$$\lambda_{k,est} = n_0 / 20$$

$$\mu_{k,est} = (\gamma_k^{(1)} + \gamma_k^{(2)}) / 2$$

$$\sigma_{k,est}^2 = n_1 (\gamma_k^{(1)} - \gamma_k^{(2)})^2 / 2$$

Donde  $n_0$  representa el número de muestras individuales que fueron identificadas por la prueba como negativas en contenido de yodo,  $n_1$  representa el número de muestras identificadas como positivas, y  $\gamma_k^{(1)}$  y  $\gamma_k^{(2)}$  de nuevo representan la cantidad de yodo reportada por el laboratorio, pero ahora de las dos muestras combinadas a partir solo de las muestras individuales positivas.

#### • Método 2

Es posible que en cada comunidad los productores de sal que la fortifican, lo hagan por medio de procesos similares, con los cuales todos obtienen una variabilidad similar de contenido de yodo en sus unidades, aunque de hecho hayan establecido diferentes niveles de contenido de yodo. Por lo que tomando muestras directamente de las plantas de producción de sal, podría estimarse el parámetro:  $\sigma_k^2$ , esto es independientemente del nivel de contenido de yodo de cada productor.



Luego todas las muestras individuales de una escuela pueden ser utilizadas para obtener 2 muestras combinadas, las cuales deben ser analizadas en el laboratorio para determinar su contenido de yodo. Digamos los valores:  $\gamma_k^{(1)}$  y  $\gamma_k^{(2)}$ , para la  $k$ -ésima comunidad. Con estos valores se pueden obtener las siguientes estimaciones.

$$v_{k,est} = (\gamma_k^{(1)} + \gamma_k^{(2)}) / 2$$

$$\tau_{k,est}^2 = 5 (\gamma_k^{(1)} - \gamma_k^{(2)})^2$$

Las cuales pueden ser utilizadas para plantear el siguiente sistema de 2 ecuaciones con las 2 incógnitas:

$\lambda_k$  y  $\mu_k$ .

$$v_{k,est} = (1 - \lambda_k) \mu_k$$

$$\tau_{k,est}^2 = (1 - \lambda_k) [\lambda_k \mu_k^2 + s_k^2]$$

Donde  $s_k^2$  representa la estimación del parámetro:  $\sigma_k^2$ , calculada en base a muestras tomadas directamente de las plantas de producción.

Claramente resolviendo este sistema de ecuaciones puede obtenerse –en forma indirecta– estimaciones de los parámetros:  $\lambda_k$  y  $\mu_k$ . Para resolver este sistema es conveniente calcular el siguiente coeficiente de variación.

$$c_k = \tau_{k,est}^2 / v_{k,est}$$

Por lo que a partir de las ecuaciones anteriores y con este coeficiente, podemos escribir la ecuación siguiente.

$$\mu_k^2 - (c_k - v_{k,est}) \mu_k + s_k^2 = 0$$

La que puede resolverse algebraicamente para  $\mu_k$ , y luego su solución puede usarse para calcular  $\lambda_k$ , como sigue.

$$\mu_k = - [(c_k - v_{k,est}) \pm \sqrt{(c_k - v_{k,est})^2 - 4s_k^2}]$$

$$\lambda_k = 1 - (v_{k,est} / \mu_k)$$

En la Tabla 4 presentamos los valores reportados por el laboratorio de las muestras combinadas de 5 escuelas encuestadas en 1999, junto con las estimaciones de sus parámetros  $\mu_k$  y  $\gamma_k$  obtenidas aplicando las fórmulas recién descritas. Para estos cálculos hemos adoptado en forma arbitraria los siguientes 4 diferentes valores de  $s_k^2 = 25, 100, 225$  y  $400$ . Esto solo para ilustrar la aplicación de las fórmulas.

Debemos notar que con la primera fórmula en general obtenemos 2 diferentes soluciones para  $\mu_k$ , de las cuales solo una corresponde al promedio de la cantidad de yodo contenida en las muestras individuales de sal. Esta solución resulta ser la correspondiente al signo + del radical. Luego con este valor de  $\mu_{k,est}$ , queda también determinada una sola solución para estimar el parámetro  $\lambda_k$ .

Además, notemos que para algunos valores específicos de  $s_k^2$ ,  $\gamma_k^{(1)}$  y  $\gamma_k^{(2)}$ , la primera ecuación podría no tener una solución en los números reales. Como es el caso de la escuela con  $k = 5$  de la misma Tabla 4, donde la poca variabilidad observada entre los valores reportados:  $\gamma_k^{(1)} = 17.2$  y  $\gamma_k^{(2)} = 13.9$ , esta es  $\tau_{k,est}^2 = 54.45$ , no está de acuerdo a los supuestos valores de  $s_k^2 = 100, 225$  y  $400$ .

**Tabla 4**Estimaciones ilustrativas de los parámetros  $\mu_k$  y  $\lambda_k$ .

$k$	$S^2_k$	$\gamma_k^{(1)}$	$\gamma_k^{(2)}$	$v_{k,est}$	$\tau^2_{k,est}$	$\mu_k$	$\lambda_k$
1	25	1.1	6.3	3.70	135.20	39.61	0.91
2	25	13.8	27.6	20.70	952.20	66.32	0.69
3	25	15.1	34.7	24.90	1,920.80	101.79	0.76
4	25	14.7	47.0	30.85	5,216.45	199.82	0.85
5	25	17.2	13.9	15.55	54.45	17.63	0.12
1	100	1.1	6.3	3.70	135.20	37.58	0.90
2	100	13.8	27.6	20.70	952.20	65.17	0.68
3	100	15.1	34.7	24.90	1,920.80	101.05	0.75
4	100	14.7	47.0	30.85	5,216.45	199.44	0.85
5	100	17.2	13.9	15.55	54.45	#NUM!	#NUM!
1	225	1.1	6.3	3.70	135.20	33.53	0.89
2	225	13.8	27.6	20.70	952.20	63.14	0.67
3	225	15.1	34.7	24.90	1,920.80	99.79	0.75
4	225	14.7	47.0	30.85	5,216.45	198.81	0.84
5	225	17.2	13.9	15.55	54.45	#NUM!	#NUM!
1	400	1.1	6.3	3.70	135.20	22.32	0.83
2	400	13.8	27.6	20.70	952.20	60.04	0.66
3	400	15.1	34.7	24.90	1,920.80	97.96	0.75
4	400	14.7	47.0	30.85	5,216.45	197.92	0.84
5	400	17.2	13.9	15.55	54.45	#NUM!	#NUM!





Roberto A. Molina Cruz  
rmolina@uvg.edu.gt

Departamento de  
Matemática de la Facultad de  
Ciencias y Humanidades  
de la Universidad del Valle de  
Guatemala

## Bibliografía

Congreso de la República de Guatemala (1995). Ley General de Enriquecimiento de Alimentos y sus Reglamentos, Decreto 44-92. Guatemala.

Molina, R., D. Fitch, y J. Lepkowski. 2001. Monitoring Salt Iodination in Developing Countries. Proceedings of the American Statistical Association. Joint Statistical Meeting, Atlanta, Georgia August 5-9. ISBN 1-931586-13-6.

UNICEF y Ministerio de Educación (1999). Escuelas Centinelas Micronutrientes. UNICEF, Guatemala.