

# Agrupamiento óptimo para una muestra con distribución normal y parámetros desconocidos

Iván Mauricio Pérez Laguna

Universidad Politécnica de Texcoco. Estado de México

---

## Resumen

En el presente artículo consideramos el caso de una muestra agrupada con función de densidad de probabilidad normal con media y desviación estándar desconocidas y proponemos los puntos de corte para formar el agrupamiento óptimo, definido como aquel que maximiza la información relativa asintótica de Fisher. Además, se presentan las expresiones de los estimadores de máxima verosimilitud y un método iterativo para su obtención. Para analizar los puntos de corte y de las expresiones obtenidas, se utilizaron la eficiencia del diseño de muestreo y el error cuadrado medio.

*Palabras clave:* Agrupamiento, información relativa asintótica, muestra agrupada, puntos de corte, varianza asintótica.

*Clasificación AMS:* 62D05, 62N02.

## Optimal spacing for a sample with normal distribution and unknown parameters

---

### Abstract

In the present paper we consider the case of a grouped sample with normal probability density function with mean and standard deviation unknown and we propose the cutoffs to form the optimal grouping, defined as the one that maximizes the asymptotic Fisher information. In addition, the equations of the maximum likelihood estimators and an iterative method to obtain them are presented. To analyze the cutoffs and the expressions obtained, the efficiency of the sample design and the mean square errors were used.

*Keywords:* Grouping, relative asymptotic information, grouped sample, cutoffs, asymptotic variance.

*AMS Classification:* 62D05, 62N02.

## 1. Introducción

Tradicionalmente la muestra de una población se compone de observaciones aleatorias individuales obtenidas bajo las reglas del muestreo aleatorio. Sin embargo, en áreas como el control de calidad, medicina clínica y epidemiología, las observaciones deben agruparse dentro de intervalos donde la eficiencia del estudio dependerá de cuales sean y como se elijan los puntos que forman los intervalos (Park, 2006). En la práctica, es usual tratar las muestras agrupadas como si no lo estuvieran. Éste procedimiento representa una aproximación que frecuentemente presenta errores sistemáticos considerables (Kulldorff, 1962). Al carecer de tablas o procedimientos analíticos para establecer los puntos de corte, es común que los investigadores establezcan estos intervalos de forma intuitiva o por experiencia. Además, los puntos de corte dependen de los parámetros de la distribución (la mayoría de las veces desconocidos) y, por tanto, contar con información acerca de los parámetros es un requisito indispensable para determinar los intervalos de agrupamiento. En sondeos de opinión, estudios de mercado y censos, existen problemas en la recolección de la información cuando, debido al tipo de preguntas el entrevistado siente que se está invadiendo su intimidad, por ejemplo: el total al que ascienden sus ingresos, el número de relaciones extramaritales que ha mantenido, una opción es ofrecer al entrevistado intervalos en los cuales pudiera colocar su respuesta probable. A su vez, es usual analizar los resultados de las encuestas sobre ingreso y riqueza, agrupando los datos en intervalos (Davies et al., 1989). Ante esta cuestión, el investigador pudiera preguntarse 1) ¿cuántos intervalos debe elegir? 2) ¿Cuáles deben ser los límites o puntos de corte? para que la información extraída sea máxima.

Kulldorff (1962) propone el agrupamiento óptimo y los estimadores de máxima verosimilitud para una muestra obtenida a partir de las distribuciones exponencial y normal cuando sólo uno de los parámetros es desconocido, definiendo como agrupamiento óptimo a aquel que maximiza la información que se puede obtener de la muestra debido a los puntos o intervalos elegidos; es decir, el agrupamiento óptimo es aquel que minimiza la varianza asintótica. Ogawa (1977) utilizó los cuantiles muestrales para obtener los mejores estimadores asintóticamente insesgados. El agrupamiento óptimo es el que maximiza la eficiencia relativa asintótica (el determinante de la matriz de información de Fisher) para una muestra con función de distribución normal cuando la media y la desviación estándar son desconocidas.

## 2. Preliminares

Considérese una población en la que se estudia una característica  $X$  que sigue una distribución normal con función de densidad dada por:

$$f_X(x_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}; \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad x_i \in \mathbb{R}, \quad (i = 0, 1, \dots, m) \quad [1]$$

De ella se obtiene una muestra de  $n$  observaciones independientes. Debido a las condiciones del estudio, el dominio de  $X$  se divide en  $m$  intervalos con la restricción

$m \geq 2$ . Sean  $h_i (i = 0, 1, \dots, m)$  los puntos donde se realizan los cortes al dominio de la función tales que  $h_i = x_i$  y  $h_{i-1} < h_i$ . Cada par ordenado  $(h_{i-1}, h_i]$  es definido como el  $i$ -ésimo intervalo, donde  $h_0$  es el límite inferior de la distribución tal que  $F_X(h_0; \theta) = 0$  y  $h_m$  es el límite superior de la distribución tal que  $F_X(h_m; \theta) = 1$ .

De esta forma [1] en términos de los puntos de corte puede expresarse como:

$$f_X(h_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{h_i - \mu}{\sigma}\right)^2}; \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad h_i \in \mathbb{R}, \quad (i = 0, 1, \dots, m) \quad [2]$$

Sea  $\xi_i = (h_i - \mu)/\sigma$  ( $i = 0, 1, \dots, m$ ) entonces:

$$f_X(h_i; \mu, \sigma) = \phi(\xi_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi_i^2} \quad \text{y} \quad F_X(h_i; \mu, \sigma) = \Phi(\xi_i) = \int_{-\infty}^{\xi_i} \phi(z_i) dz_i \quad [3]$$

Con las condiciones expuestas, la probabilidad de que una observación caiga en el  $i$ -ésimo intervalo está dada por:

$$\pi_i = P(h_{i-1} < X_j \leq h_i) = F_X(h_i; \theta) - F_X(h_{i-1}; \theta) = \Phi(\xi_i) - \Phi(\xi_{i-1}) \quad [4]$$

donde,  $\pi_i \geq 0$ ,  $\sum_{i=1}^m \pi_i = 1$ ,  $n_i$  ( $i = 1, 2, \dots, m$ ) es el número de observaciones en el  $i$ -ésimo intervalo y  $n = \sum_{i=1}^m n_i$

Además, el logaritmo neperiano de la verosimilitud total es:

$$\ln L(\theta, n_i) = \ln(\pi_1^{n_1} \pi_2^{n_2} \dots \pi_m^{n_m}) = \ln \prod_{i=1}^m \pi_i^{n_i} = \sum_{i=1}^m n_i \ln \pi_i = \sum_{i=1}^m n_i \ln [\Phi(\xi_i) - \Phi(\xi_{i-1})] \quad [5]$$

### 3. Estimadores de máxima verosimilitud

El estimador de máxima verosimilitud (EMV)  $\hat{\theta}$  de  $\theta$  es definido cómo:

$$\hat{\theta} = MLE(\theta) = \frac{\partial \ln L(\theta; n_i)}{\partial \theta} = \sum_{i=1}^m n_i \frac{\partial \ln \pi_i}{\partial \theta} = \sum_{i=1}^m n_i \frac{\partial \ln [\Phi(\xi_i) - \Phi(\xi_{i-1})]}{\partial [\mu, \sigma]^T} = 0 \quad [6]$$

Derivando parcialmente [6] respecto a  $\mu$  y  $\sigma$  las siguientes expresiones de los EMV son obtenidas:

$$\sum_{i=1}^m n_i \frac{\phi(\xi_i) - \phi(\xi_{i-1})}{\Phi(\xi_i) - \Phi(\xi_{i-1})} = 0$$

$$\sum_{i=1}^m n_i \frac{(\xi_i)\phi(\xi_i) - (\xi_{i-1})\phi(\xi_{i-1})}{\Phi(\xi_i) - \Phi(\xi_{i-1})} = 0$$
[7]

#### 4. La varianza asintótica

La varianza de la distribución asintótica del estimador de máxima verosimilitud  $\hat{\theta}$  está dada (Kulldorff, 1962) por:

$$\text{AsVar}(\hat{\theta}) = E \left[ \left( \frac{\partial \ln \pi_i}{\partial \theta} \right)^2 \right]^{-1} = \left[ n \sum_{i=1}^m \pi_i \left( \frac{\partial \ln \pi_i}{\partial \theta} \right)^2 \right]^{-1} =$$

$$= \left( -n \sum_{i=1}^m \Phi(\xi_i) - \Phi(\xi_{i-1}) \frac{\partial^2 \ln [\Phi(\xi_i) - \Phi(\xi_{i-1})]}{\partial([\mu, \sigma]^t)^2} \right)^{-1}$$
[8]

Derivando a [8] dos veces con respecto  $\mu$  y  $\sigma$  la expresión de la varianza asintótica es obtenida:

$$\text{AsVar}(\hat{\theta}) = \frac{\sigma^2}{n} \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix}$$
[9]

donde,

$$a_1 = \sum_{i=1}^m \left\{ \left[ \xi_i \phi(\xi_i) - \xi_{i-1} \phi(\xi_{i-1}) \right] + \frac{[\phi(\xi_i) - \phi(\xi_{i-1})]^2}{\Phi(\xi_i) - \Phi(\xi_{i-1})} \right\}$$
[9.1]

$$a_3 = \sum_{i=1}^m \left\{ \left[ \xi_i \phi(\xi_i) [\xi_i^2 - 1] - \xi_{i-1} \phi(\xi_{i-1}) [\xi_{i-1}^2 - 1] \right] + \frac{[\xi_i \phi(\xi_i) - \xi_{i-1} \phi(\xi_{i-1})]^2}{[\Phi(\xi_i) - \Phi(\xi_{i-1})]} \right\}$$
[9.2]

$$a_2 = \sum_{i=1}^m \left\{ \phi(\xi_i) - \phi(\xi_{i-1}) - \left[ \xi_i^2 \phi(\xi_i) - \xi_{i-1}^2 \phi(\xi_{i-1}) \right] - \frac{[\phi(\xi_i) - \phi(\xi_{i-1})][\xi_i \phi(\xi_i) - \xi_{i-1} \phi(\xi_{i-1})]}{[\Phi(\xi_i) - \Phi(\xi_{i-1})]} \right\}$$
[9.3]

El valor numérico de [9] es la información relativa asintótica (Kulldorff, 1962). La matriz en [9] es definida negativa con determinante (Ogawa, 1977):

$$\Delta = a_1 \times a_3 - [a_2]^2 \quad [10]$$

## 5. Agrupamiento óptimo

El agrupamiento óptimo está formado por los puntos de corte que hacen que la varianza [9] sea la menor posible. Minimizar [9] equivale a maximizar la matriz de esta expresión; es decir, maximizar [10] (Ogawa, 1977). En el Cuadro 1 están los puntos de corte óptimos  $\xi_i = (h_i - \mu)/\sigma (i = 1, 2, \dots, 11)$  obtenidos al maximizar [10] y la información relativa asintótica óptima derivada al substituir las  $\xi_i$  en [9].

Cuadro 1

### Puntos de corte óptimos y la información relativa asintótica

$m$	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$\xi_6$	$\xi_7$	$\xi_8$	$\xi_9$	$\xi_{10}$	Información Relativa Asintótica
2	-1,171										0,195
3	-1,107	1,107									0,822
4	-1,376	0,011	1,384								1,115
5	-1,691	-0,687	0,687	1,691							1,374
6	-1,875	-0,994	0	0,994	1,875						1,519
7	-2,050	-1,260	-0,489	0,489	1,260	2,050					1,627
8	-2,183	-1,449	-0,783	0	0,783	1,449	2,183				1,701
9	-2,303	-1,613	-1,017	-0,380	0,380	1,017	1,613	2,303			1,756
10	-2,403	-1,747	-1,198	-0,646	0	0,646	1,198	1,747	2,403		1,797
11	-2,493	-1,865	-1,352	-0,857	-0,312	0,312	0,857	1,352	1,865	2,493	1,828

El agrupamiento óptimo que maximiza la eficiencia relativa asintótica es simétrico (Ogawa, 1977). Los puntos de corte óptimos  $\xi_i$  (Cuadro 1) confirman este resultado; las  $\xi_i$  son simétricas cuando  $m$  es impar, pero cuando  $m$  es par si disminuye la precisión prácticamente se tiene simetría. Se puede observar (Cuadro 1) que la información relativa asintótica aumenta conforme aumenta el número de puntos de corte; sin embargo, a partir de 8 puntos este aumento no es significativo.

## 6. Obtención de los estimadores de máxima verosimilitud

Para obtener los estimados de máxima verosimilitud, es necesario resolver simultáneamente el sistema de ecuaciones [7]. Este sistema depende de dos parámetros desconocidos,  $\mu$  y  $\sigma$ , lo que hace difícil obtener una solución cerrada. Las expresiones presentadas en este artículo se analizan simulando una muestra de tamaño 200, con una media de 50 y una desviación estándar de 10, permitiendo proponer un método iterativo que ayudará a obtener una solución en casos prácticos. Se analizan los resultados numéricos de las expresiones propuestas usando la eficiencia del diseño de muestreo (*deff*) (Kish, 1965), un cociente

de varianzas que permite comparar el agrupamiento del método iterativo contra el agrupamiento óptimo (Cuadro 1):

$$def_{p(s)}(\hat{\theta}) = \frac{V_{a(a)}(\hat{\theta})}{V_{a(b)}(\hat{\theta})} \quad [11]$$

donde,  $def_{p(s)}(\hat{\theta})$  es el indicador de la eficiencia del diseño muestral,  $V_{a(a)}(\hat{\theta})$  es la varianza del estimador con el método a comparar,  $V_{a(b)}(\hat{\theta})$  es la varianza del estimador con el que se compara.

Y el error cuadrado medio (ECM) para comparar a los EMV contra los parámetros de la muestra:

$$ECM(\theta) = E\left\{\left[\hat{\theta} - \theta\right]^2\right\} = \left[\ell(\hat{\theta}) - \theta\right]^2 \quad [12]$$

donde,  $\hat{\theta}$  es el  $MLE(\theta)$ ,  $\ell(\hat{\theta})$  su valor numérico y  $\theta' = [\mu, \sigma]$ .

## 7. El método iterativo

Partiendo de una muestra con función de densidad de probabilidad normal, haga un agrupamiento inicial utilizando la relación numérica:

$$\tilde{h}_i = y_{(1)} + i \times \frac{y_{(n)} - y_{(1)}}{m} \quad [13]$$

donde,  $m$  es el número de intervalos,  $y_{(1)} = \text{menor}\{y_i\}_{i=1}^n$ ,  $y_{(n)} = \text{mayor}\{y_i\}_{i=1}^n$ ,  $\tilde{h}_i$  ( $i = 1, 2, \dots, m$ ) se definirá como el  $i$ -ésimo punto de corte semilla.

Como estimadores iniciales para la media y la desviación estándar se puede utilizar las expresiones para la media y la desviación estándar muestrales para datos agrupados:

$$\tilde{\mu} = \frac{\sum_{i=1}^m n_i u_i}{\sum_{i=1}^m n_i} \quad [14]$$

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^m n_i u_i^2}{\sum_{i=1}^m n_i} - (\tilde{\mu}^2)$$

donde,  $u_i = \frac{h_i + h_{i+1}}{2}$  es el valor medio del intervalo o clase ( $i=2,3,\dots,m-1$ ), y  $n_i$  es el número de observaciones en el  $i$ -ésimo intervalo. Las expresiones en [14] son estimadores semilla.

De esta manera, la metodología a seguir para encontrar las estimaciones de máxima verosimilitud es la siguiente:

- 1) Se obtienen los puntos de corte semilla con la expresión [13] y se realiza un agrupamiento inicial calculando la información relativa asintótica semilla, substituyendo los valores  $\xi_i = (\tilde{h}_i - \tilde{\mu}) / \tilde{\sigma}$  en [9].
- 2) Substituyendo las  $\xi_i$  semilla en el sistema [7], se obtienen los estimadores de máxima verosimilitud iniciales  $\hat{\mu}^0$  y  $\hat{\sigma}^0$ ; luego, se obtienen los puntos de corte iniciales al substituir estos estimados iniciales en  $\xi_i = (\tilde{h}_i - \hat{\mu}^0) / \hat{\sigma}^0$  para calcular la información relativa asintótica inicial con la expresión [9].
- 3) Con los puntos de corte iniciales  $h_i = \hat{\sigma}^0 \xi_i + \hat{\mu}^0$ , donde  $\xi_i$  el  $i$ -ésimo punto de corte óptimo del Cuadro 1, se realiza un *nuevo agrupamiento* que será considerado como agrupamiento óptimo.
- 4) Con la información del agrupamiento óptimo y resolviendo [7] es posible obtener los estimados de máxima verosimilitud óptimos  $\hat{\mu}$ ,  $\hat{\sigma}$  y de manera análoga su información relativa asintótica óptima substituyendo  $\xi_i = (h_i - \hat{\mu}) / \hat{\sigma}$  en [9].

En los cuadros 2 y 3 se presentan los resultados obtenidos aplicando el método iterativo a la muestra simulada.

Cuadro 2

**Estimados semillas, iniciales y óptimos para una muestra de tamaño doscientos obtenida de una población normal con parámetros  $m=50$  y  $s=10$**

$m$	<i>Semillas</i>		<i>Iniciales</i>		<i>Óptimos</i>		<i>MSE</i>	<i>MSE</i>	<i>MSE</i>	<i>MSE</i>	<i>MSE</i>	<i>MSE</i>
	$\tilde{\mu}$	$\tilde{\sigma}$	$\hat{\mu}^0$	$\hat{\sigma}^0$	$\hat{\mu}$	$\hat{\sigma}$	$(\tilde{\mu})$	$(\hat{\mu}^0)$	$(\hat{\mu})$	$(\tilde{\sigma})$	$(\hat{\sigma}^0)$	$(\hat{\sigma})$
2	50,713	14,345	50,959	4,412	48,376	4,923	0,508	0,920	2,639	18,882	31,232	25,778
3	50,260	10,928	50,224	9,495	49,897	10,033	0,068	0,050	0,011	0,862	0,255	0,001
4	50,478	9,9710	50,463	9,045	50,247	9,850	0,228	0,214	0,061	0,001	0,911	0,022
5	50,730	10,262	50,723	9,701	50,322	9,770	0,533	0,522	0,104	0,069	0,089	0,053
6	49,886	11,405	50,421	9,597	50,241	9,708	0,013	0,178	0,058	1,973	0,163	0,086
7	50,409	10,097	50,406	9,815	50,351	9,920	0,167	0,165	0,124	0,009	0,034	0,006
8	50,363	9,896	50,362	9,675	50,357	9,806	0,132	0,131	0,127	0,011	0,105	0,038
9	50,518	9,805	50,516	9,630	50,379	9,610	0,268	0,266	0,144	0,038	0,137	0,152
10	50,442	9,998	50,440	9,859	50,488	10,100	0,195	0,194	0,238	0	0,020	0,010
11	50,405	10,104	50,404	9,990	50,587	9,758	0,164	0,164	0,345	0,011	0	0,059

En el Cuadro 2 puede verse que el error cuadrado medio de los estimados de máxima verosimilitud óptimos  $(\hat{\mu}, \hat{\sigma})$  tiene un valor menor que el de los otros estimados (semillas e iniciales) tanto para la media como para la desviación estándar. Esto indica que el valor numérico obtenido por los EMV propuestos en este artículo es más cercano al valor de los parámetros de la muestra simulada.

Cuadro 3.

**Información relativa asintótica y eficiencia del muestreo de los estimadores semilla, iniciales y óptimos para una muestra de tamaño doscientos obtenida de una población normal con parámetros**

<i>m</i>	<i>Óptima</i>	<i>Semilla</i> $(\tilde{\mu}, \tilde{\sigma})$	<i>Inicial</i> $(\hat{\mu}^0, \hat{\sigma}^0)$	<i>Óptimos</i> $(\hat{\mu}, \hat{\sigma})$	<i>deff</i>				
					$(\tilde{\mu}, \tilde{\sigma})$	$(\hat{\mu}^0, \hat{\sigma}^0)$	$(\hat{\mu}, \hat{\sigma})$	$\frac{(\hat{\mu}, \hat{\sigma})}{(\hat{\mu}^0, \hat{\sigma}^0)}$	$\frac{(\hat{\mu}, \hat{\sigma})}{(\tilde{\mu}, \tilde{\sigma})}$
2	0,195	0,497	0,001	0,075	0,393	255,49	2,601	98,227	0,151
3	0,822	0,947	0,898	0,807	0,869	0,916	1,018	0,899	0,853
4	1,115	1,194	1,123	1,091	0,934	0,993	1,022	0,972	0,914
5	1,374	1,408	1,393	1,364	0,976	0,987	1,007	0,980	0,969
6	1,519	1,534	1,528	1,511	0,990	0,994	1,006	0,989	0,985
7	1,628	1,621	1,623	1,62	1,004	1,003	1,004	0,998	1,000
8	1,701	1,685	1,688	1,694	1,009	1,008	1,004	1,004	1,005
9	1,756	1,733	1,736	1,751	1,013	1,011	1,003	1,008	1,010
10	1,797	1,76	1,764	1,791	1,021	1,018	1,003	1,015	1,018
11	1,828	1,779	1,784	1,824	1,028	1,025	1,002	1,022	1,025

Los valores del indicador *deff* de los estimados óptimos  $(\hat{\mu}, \hat{\sigma})$  son muy cercanos a 1 cuando *m* es menor o igual a 5 (Cuadro 3). Esto indica que la información que se puede obtener de la muestra debido al agrupamiento derivado de los estimadores óptimos propuestos es más cercana a la información óptima. Sin embargo, cuando *m* es menor o igual a 6 todos los indicadores toman valores cercanos a 1, lo cual indica que para todos los estimados la información obtenida a partir de 6 puntos de corte es la misma. Es decir, no hay pérdida de información debido al agrupamiento elegido.

## 8. Conclusiones

En este artículo se propusieron los puntos de corte para formar un agrupamiento óptimo de tal forma que la varianza sea mínima. Las expresiones de los estimadores de máxima verosimilitud fueron analizadas junto con un método iterativo para su obtención. Los puntos de corte propuestos pueden utilizarse para establecer un agrupamiento del cual la información que se obtendrá será máxima. Si se utilizan los estimadores junto con los puntos de corte propuestos, es posible conocer la media y la desviación estándar de la muestra. Se recomienda utilizar el Cuadro 1 para agrupar una muestra en casos prácticos.



## Referencias

---

- DAVIES, JB.; SHORROCKS, AF. (1989). «Optimal grouping of income and wealth data. Colchester», *Journal of Econometrics*, 42: 97–108.
- KULLDORFF, G. (1962): «Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples». *New York, Wiley*.
- KISH, L. (1965): «Survey Sampling». *New York, J. Wiley. Wiley Classic Library Edition*.
- OGAWA, J. (1977): «Optimal spacings for the simultaneous estimation of the location and scale parameters of a normal distribution based on selected two sample quantiles». *Journal of Statistical Planning and Inference*, 1, 61 – 72.
- PARK, S. (2006): «Conditional optimal spacing in exponential distribution». *Lifetime Data Analysis* 12, 523 – 530.