Revista Colombiana de Estadística
No. 31 y No. 32

UNA ESTADISTICA LINEAL DE RANGOS BASADA EN DESARROLLO BINARIO PARA ANALISIS DE LOCALIZACION

J. HUMBERTO MAYORGA A.

Profesor Asociado
Universidad Nacional de Colombia

RESUMEN. **Este artículo propone una estadística lineal de rangos para el análisis de localización de dos poblaciones , basada en la representación binaria de enteros. Su naturaleza particular le permite poseer una densidad con una expresión muy sencilla y por lo cual su distribución no requiere tabulación alguna.**

1. INTRODUCCIÓN

Una estadística basada en el desarrollo binario, tiene la propiedad de condensar en un único valor la información referente a un conjunto de variables dicotómicas y simultáneamente de conservar toda la información original (Mayorga 1993), hechos relevantes que hacen atrayentes desde algunos puntos de vista a este tipo de estadísticas frente a estadísticas de otra naturaleza.

Con la elección de los dígitos 0 y 1 para denotar respectivamente la pertenencia de los elementos a cada una de las dos muestras independientes y su utilización en el arreglo que describe la configuración de las rachas, puede asimilarse dicho arreglo con el desarrollo binario de un entero. La construcción de la estadística que propone este artículo, se origina en esta semejanza y se fundamenta en la condensación en un número entero, de la disposición que presenta la sucesión de rachas derivada de la muestra combinada.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Las dos muestras independientes seleccionadas respectivamente de cada una de las dos poblaciones en comparación, se denotan por $X_1, X_2, ..., X_m$, (la muestra aleatoria de tamaño m de la primera población), y por $Y_1, Y_2, ..., Y_n$ (la muestra de tamaño n de la segunda población), y sus funciones de distribución $F_X(\cdot)$ y $F_Y(\cdot)$ se asumen continuas. Reunidas las dos muestras conforman una única muestra **ordenada** de tamaño $m + n = N$, denominada muestra ordenada combinada. A partir de esta muestra aleatoria combinada se genera a su vez la muestra aleatoria $Z_1, Z_2, ..., Z_N$, sucesión de variables aleatorias dicotómicas definidas de la siguiente manera:

Si el i-ésimo elemento de la muestra combinada pertenece a la **primera** muestra, Z_i asume el valor 1; por el contrario cuando dicho elemento pertenece a la **segunda** muestra, Z_i asume el valor 0, para $i = 1, 2, ..., N$.

1.1 Definición de la Estadística.

Las variables dicotómicas Z_i, $i = 1, 2, ..., N$, desempeñan únicamente el papel indicador de la pertenencia de un elemento muestral a una de las dos muestras. Sin embargo una realización específica de dichas variables $z_1, z_2, ..., z_N$, vista como una sucesión particular de rachas puede manifestar alguna disposición de los elementos de la muestra combinada como indicativo de diferencias acentuadas de localización entre las dos poblaciones en comparación.

La estadística,

$$I_{(m,n)} = \sum_{i=1}^{N} 2^{N-i} Z_i$$

asigna a cada sucesión particular de rachas, expresada por una sucesión de dígitos 0 y 1, un número entero. La interpretación de esta sucesión como representación binaria del entero asignado, permite establecer una correspondencia biunívoca entre las sucesiones de rachas y el recorrido de la estadística que garantiza el poder recuperar plenamente la estructura de una sucesión particular de rachas a partir de un valor específico de la estadística.

El valor del entero asignado depende en consecuencia de la ubicación de las muestras reflejada en la localización de los dígitos por supuesto. En el caso particular que todos los elementos de la primera muestra se ubiquen en los últimos m lugares de la muestra combinada (por lo tanto los primeros n corresponden a la segunda muestra), $I_{(m,n)}$ presenta su mínimo valor.

En efecto, al ubicarse totalmente los elementos de la primera muestra a la derecha de los de la segunda muestra, significa que $z_i = 0$ para $i = 1, 2, ..., n$ y que $z_i = 1$ para $i = n+1, n+2, ..., N$, y en consecuencia el valor específico de $I_{(m,n)}$ en estas circunstancias es:

$$\begin{aligned} I_{(m,n)} &= \sum_{i=1}^{N} 2^{N-i} Z_i = \sum_{i=1}^{n} 2^{N-i} 0 + \sum_{i=n+1}^{N} 2^{N-i} \\ &= \sum_{j=0}^{m-1} 2^{(N-n)-j} = \sum_{j=1}^{m} 2^{m-j} = 2^m - 1 \end{aligned}$$

Cuando la ubicación es totalmente contraria a la anterior,$I_{(m,n)}$ alcanza su máximo valor; este valor se obtiene teniendo en cuenta que:

$$\begin{aligned} I_{(m,n)} &= \sum_{i=1}^{N} 2^{N-i} Z_i = \sum_{i=1}^{m} 2^{N-i} + \sum_{i=m+1}^{N} 2^{N-i} 0 \\ &= \sum_{i=1}^{m} 2^{N-i} = 2^N - 2^n \end{aligned}$$

Valores pequeños de la estadística corresponden a manifestaciones de tendencia de localización de la primera muestra a la derecha de la segunda muestra, valores intermedios muestran la no distinción de la localización de una y otra muestra y los valores grandes indicativos de tendencia de localización de la primera muestra a la izquierda de la segunda muestra.

1.2. Distribución de $I_{(m,n)}$.

Como a cada disposición de la rachas configuradas por m símbolos "1" y n símbolos "0", le corresponde un único entero, el número de valores de la estadística $I_{(m,n)}$ coincide con el número de disposiciones distintas de las rachas; esto es, como el número de arreglos posibles de m unos y n ceros es $\frac{N!}{n!m!} = \Omega(m,n)$, el recorrido de la estadística $I_{(m,n)}$, tendrá por supuesto $\Omega(m,n)$ valores, recorrido denotado como $R_{I_{(m,n)}} = \{l_1, l_2, ..., l_{\Omega(m,n)}\}$.

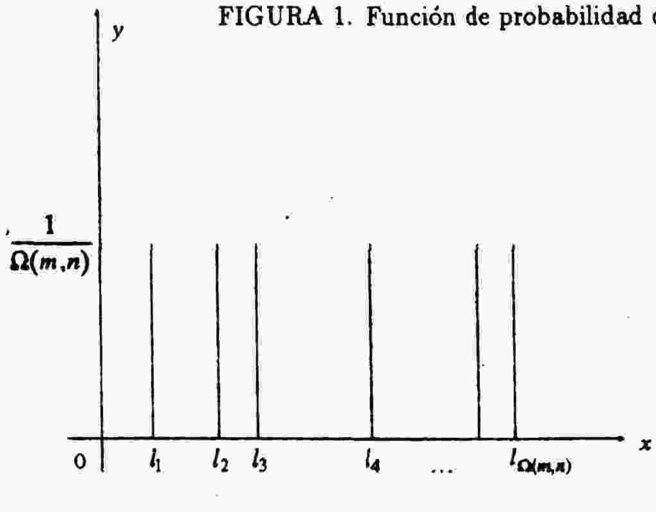
Bajo el supuesto de total aleatoriedad, cada arreglo es equiprobable, por lo tanto,

$$P[I_{(m,n)} = l_k] = \begin{cases} \frac{1}{\Omega(m,n)} & si \;\; l_k \in R_{I_{(m,n)}} \\ 0 & si \;\; l_k \notin R_{I_{(m,n)}} \end{cases}$$

$$P[I_{(m,n)} \leq l_k] = \frac{k}{\Omega(m,n)} \quad k = 1, 2, ..., \Omega(m,n)$$

Aun cuando cada uno de los valores de $I_{(m,n)}$ tiene igual probabilidad, es preciso señalar que la estadística no tiene la distribución uniforme discreta comunmente conocida; el recorrido especial de ella determina la diferencia, como lo sugiere la Figura 1.

FIGURA 1. Función de probabilidad de $I_{(m,n)}$



La estadística $I_{(m,n)}$ pertenece a la clase de estadísticas denominada **Estadísticas lineales de rangos**, (*Gibbons 1971*) clase que agrupa a las estadísticas de la forma,

$$\sum_{i=1}^{N} a_i Z_i$$

donde los coeficientes a_i, son valores predeterminados; basta considerar $a_i = 2^{N-i}$, $i = 1, 2, ..., N$, para identificar a $I_{(m,n)}$ como miembro de dicha clase.

A partir de las expresiones generales para el valor esperado y varianza de las estadísticas lineales de rangos, bajo la hipótesis nula $H_0 : F_X(x) = F_Y(x)$, para todo x, se puede determinar que:

$$\mathbb{E}[I_{(m,n)}] = \frac{m}{N}\sum_{i=1}^{N} 2^{N-i} = \frac{m}{N}(2^N - 1)$$
$$\mathbb{V}[I_{(m,n)}] = \frac{mn}{N^2(N-1)}\left\{\frac{N}{3}(4^N - 1) - (2^N - 1)^2\right\}$$

2. Analisis de Localización con Base en la estadística $I_{(m,n)}$.

2.1. Qué identifica la estadística ?

Asimilar la sucesión de rachas derivada de los valores $z_1, z_2, ..., z_N$, como la representación binaria de un número entero, como ya se ha dicho, además de permitir su condensación en un único valor (el entero correspondiente) que conserva inalterada su disposición, abre posibilidades de diagnóstico sobre la naturaleza de la localización de las dos poblaciones cotejadas. Por medio de un ejemplo, poco realista a la luz de los dos tamaños muestrales, pero elegido por su simplicidad, se inducen las ideas iniciales del propósito de utilizar la estadística $I_{(m,n)}$ como estadística central en el análisis de localización.

Una muestra de tamaño 3 correspondiente a la primera población y una muestra de tamaño 4 proveniente de la segunda población, pueden originar $\Omega(3,4) = 35$ disposiciones diferentes de sucesiones de rachas de tres simbolos "1" y cuatro simbolos "0", como a continuación se listan en la Tabla 1, frente a ellas los enteros correspondientes a la representación binaria, es decir los valores de la estadística en consideración y adicionalmente el orden que representa cada valor.

La primera disposición 0000111 representación binaria de 7, mínimo valor de la estadística particular $I_{(3,4)}$ corresponde a la situación extrema en la cual la muestra de la primera población se úbica plenamente a la derecha de la muestra de la segunda población; el siguiente valor 11, condensa la información correspondiente a una disposición que difiere mínimamente de la primera, pero que continuaría siendo argumento estadístico que respalda diferencias en localización de las poblaciones. A medida que se avanza en cada valor de la estadística, esa configuración inicial extrema, se va desdibujando, pasando por una zona, que no identifica a un patrón especial que corresponde a localización similar de las poblaciones, representada por los valores intermedios de la estadística, para dar paso finalmente a una nueva configuración asociada con los mayores valores, opuesta a la inicial. El máximo y último valor, obtenido de la disposición 1110000 representa por tanto otra situación extrema, antípoda de la primera disposición.

En pocas palabras, valores pequeños o valores grandes de $I_{(m,n)}$ son indicios que identifican diferencias significativas en la localización de las poblaciones, esto es, valores que constituyen argumentos estadísticos de apoyo a hipótesis que manifiestan diferencias notables de localización en algún sentido.

TABLA 1. Valores de $I_{(3,4)}$ según cada disposición de las sucesiones de rachas

z_1	z_2	z_3	z_4	z_5	z_6	z_7	$I_{(3,4)}$	J
0	0	0	0	1	1	1	7	1
0	0	0	1	0	1	1	11	2
0	0	0	1	1	0	1	13	3
0	0	0	1	1	1	0	14	4
0	0	1	0	0	1	1	19	5
0	0	1	0	1	0	1	21	6
0	0	1	0	1	1	0	22	7
0	0	1	1	0	0	1	25	8
0	0	1	1	0	1	0	26	9
0	0	1	1	1	0	0	28	10
0	1	0	0	0	1	1	35	11
0	1	0	0	1	0	1	37	12
0	1	0	0	1	1	0	38	13
0	1	0	1	0	0	1	41	14
0	1	0	1	0	1	0	42	15
0	1	0	1	1	0	0	44	16
0	1	1	0	0	0	1	49	17
0	1	1	0	0	1	0	50	18
0	1	1	0	1	0	0	52	19
0	1	1	1	0	0	0	56	20
1	0	0	0	0	1	1	67	21
1	0	0	0	1	0	1	69	22
1	0	0	0	1	1	0	70	23
1	0	0	1	0	0	1	73	24
1	0	0	1	0	1	0	74	25
1	0	0	1	1	0	0	76	26
1	0	1	0	0	0	1	81	27
1	0	1	0	0	1	0	82	28
1	0	1	0	1	0	0	84	29
1	0	1	1	0	0	0	88	30
1	1	0	0	0	0	1	97	31
1	1	0	0	0	1	0	98	32
1	1	0	0	1	0	0	100	33
1	1	0	1	0	0	0	104	34
1	1	1	0	0	0	0	112	35

2.2. Test para análisis de localización.

A partir de la formulación general,

$$H_0 : F_Y(s) = F_X(s) \quad \text{para todo} \quad s$$

$$vs. \quad H_1 : F_Y(s) = F_X(s - \theta) \quad \text{para todo } s \text{ y algún } \theta \neq 0,$$

tres sistemas de hipótesis pueden plantearse:

a. $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$

b. $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$

c. $H_0 : \theta = 0$ vs. $H_1 : \theta < 0$

Inicialmente los tests correspondientes se pueden establecer así:

γ_a :"Rechazar H_0 al nivel 100α % si el valor particular $i_{(m,n)}$ de $I_{(m,n)}$ es tal que,

$$i_{(m,n)} \leq i_{(m,n,\frac{\alpha}{2})} \quad \text{ó} \quad i_{(m,n)} \geq i_{(m,n,\frac{1-\alpha}{2})}\text{"}$$

γ_b :"Rechazar H_0 al nivel 100α % si el valor particular $i_{(m,n)}$ de $I_{(m,n)}$ es tal que,

$$i_{(m,n)} \geq i_{(m,n,1-\alpha)}\text{"}$$

γ_c :"Rechazar H_0 al nivel 100α % si el valor particular $i_{(m,n)}$ de $I_{(m,n)}$ es tal que,

$$i_{(m,n)} \leq i_{(m,n,\alpha)}\text{"}$$

siendo $i_{(m,n,\delta)} \in R_{I_{(m,n)}}$ tal que $P[I_{(m,n)} \leq i_{(m,n,\delta)}] = \delta$.

Sin embargo las propiedades de la estadística y su distribución, permiten expresar de manera equivalente dichos tests más sencillamente.

En efecto, como $i_{(m,n,\delta)} \in R_{I_{(m,n)}}, i_{(m,n,\delta)} = l_k$ para algún número natural k, $k = 1, 2, 3, ..., \Omega(m,n)$,

$$P[I_{(m,n)} \leq i_{(m,n,\delta)}] = \delta = P[I_{(m,n)} \leq l_k] = \frac{k}{\Omega(m,n)},$$

por tanto, para un valor escogido de $k, \delta = \frac{k}{\Omega(m,n)}$

Igualmente, como $i_{(m,n)} \in R_{I_{(m,n)}}, i_{(m,n)} = l_j$ para algún número natural j,

$j = 1, 2, 3, ..., \Omega(m,n)$.

De otro lado,

$$P[I_{(m,n)} \geq l_{\Omega(m,n)-(k-1)}] = \delta = P[I_{(m,n)} \leq l_k] = \frac{k}{\Omega(m,n)},$$

por supuesto $P[I_{(m,n)} < l_{\Omega(m,n)-(k-1)}] = 1 - \delta$, o sea,

$$P[I_{(m,n)} \leq l_{\Omega(m,n)-k}] = 1 - \delta,$$

es decir, si para algún $k = 1, 2, 3, ..., \Omega(m,n)$,

$$i_{(m,n,\delta)} = l_k, \text{entonces} \quad i_{(m,n,1-\delta)} = l_{\Omega(m,n)-k}$$

En consecuencia, los test pueden expresarse en la siguiente forma:

γ_a : "Rechazar H_o al nivel del 100α % si $j \leq k$ ó si $j \geq \Omega(m,n) - (k-1)$ "

siendo $\frac{\alpha}{2} = \frac{k}{\Omega(m,n)}$ para una escogencia de k, y j el subíndice correspondiente a $i_{(m,n)}$, es decir j tal que $i_{(m,n)} = l_j$.

γ_b : "Rechazar H_o al nivel del 100α % si $j \geq \Omega(m,n) - (k-1)$ "

siendo $\alpha = \frac{k}{\Omega(m,n)}$ para una escogencia de k, y j el subíndice correspondiente a $i_{(m,n)}$, es decir j tal que $i_{(m,n)} = l_j$.

γ_c : "Rechazar H_o al nivel del 100α % si $j \leq k$"

siendo $\alpha = \frac{k}{\Omega(m,n)}$ para una escogencia de k, y j el subíndice correspondiente a $i_{(m,n)}$, es decir j tal que $i_{(m,n)} = l_j$.

3. Determinación Del Subíndice Correspondiente A $i_{(m,n)}$

3.1. Consideraciones preliminares.

En primer lugar, la estadística $I_{(m,n)}$ es una aplicación 1-1 del conjunto de todas las permutaciones de m "unos" y n"ceros", en un subconjunto de números naturales, cuya sucesión de valores de su recorrido, $l_1, l_2, ...l_{\Omega(m,n)}$ está en correspondencia con una organización especial de cada una de las permutaciones sujeta a un patrón particular, que dicha sucesión permite $R_{I_{(m,n)}}$ revelar. Una partición de este recorrido en $(n+1)$ subconjuntos que se denominarán **niveles**, manifiesta esa organización especial, en la cual cada subconjunto corresponde a un grupo particular de permutaciones.

Teniendo en cuenta que cada una de las permutaciones está asociada con un único vector $z = (z_1, z_2, ..., z_N)$, el nivel 0, está conformado por el único elemento, $l_1 = 2^m - 1$, el mínimo valor de la estadística, obtenido de la permutación asociada con el vector en el cual los n primeros componentes son ceros y los m últimos componentes son unos.

El nivel 1 constituido por m elementos, corresponde a los valores de la estadística obtenidos de las permutaciones asociadas con vectores en los cuales los primeros $(n-1)$ componentes son ceros y el n-ésimo componente es un uno. El valor máximo en este nivel es el entero $(2^{m+1} - 1) - 1$.

El nivel 2 constituido por $\frac{m(m+1)}{2}$ elementos, corresponde a los valores de la estadística obtenidos de las permutaciones asociadas con vectores en los cuales los primeros $(n-2)$ componentes son ceros y el $(n-1)$- ésimo componente es un uno. El valor máximo en este nivel es el entero $(2^{m+2} - 1) - (2^2 - 1)$.

En general el nivel r constituido por $d_r(m) = \frac{(m+r-1)!}{(m-1)!r!}$ elementos,

$r = 0, 1, 2, ..., n$, corresponde a los valores de la estadística obtenidos de las permutaciones asociadas con vectores en los cuales los primeros $(n-r)$ componentes son ceros y simultáneamente el $(n-(r-1))$-ésimo componente es un uno. Para simplificar considerablemente su cómputo, $d_r(m)$ puede calcularse utilizando la expresión recurrente $d_{r+1}(m) = \left[\frac{m+r}{r+1}\right] d_r(m)$, $r = 0, 1, 2, ..., n$. El valor máximo en este nivel es el entero

$$(2^{m+r} - 1) - (2^r - 1) = 2^r[2^m - 1].$$

El número de valores que pertenecen tanto al nivel r como a los anteriores, $D_r(m) = \sum_{l=0}^{r} d_l(m)$, es simplemente $D_r(m) = \sum_{l=0}^{r} \frac{(m+l-1)!}{(m-1)!l!} = \frac{(m+r)!}{m!r!}$; por ello $D_n(m) = \frac{N!}{n!m!} = \Omega(m,n)$.

3.2. Procedimiento

Paso 1. Determinación del nivel al cual pertenece $i_{(m,n)}$.

Si el valor $i_{(m,n)}$ pertenece al nivel r, es porque él es menor o igual al máximo del nivel y a su vez es mayor que el máximo valor del nivel $r-1$. En otros términos, para algún $r = 0, 1, 2, ..., n$,

$$2^{r-1}[2^m - 1] < i_{(m,n)} \leq 2^r[2^m - 1],$$

lo cual permite establecer que $i_{(m,n)}$ pertenece al nivel r, $r = 0, 1, 2, ..., n$, si

$$2^{r-1} < \frac{i_{(m,n)}}{2^m - 1} \leq 2^r .$$

Como se deduce, a partir de un sencillo cociente entre el valor particular de la estadística y el mínimo de ella, se determina el grupo al cual pertenece. Puede en este punto concluir todo el procedimiento, si se presenta alguna de las dos situaciones siguientes: la primera en la cual $\frac{i_{(m,n)}}{2^m-1} = 2^r$, el valor es precisamente el valor máximo del nivel, en cuyo caso el subíndice de $i_{(m,n)}$ es $D_r(m)$, y en consecuencia el siguiente paso no es pertinente, y por lo tanto en términos del test, $j = D_r(m)$; la segunda cuando $1 < \frac{i_{(m,n)}}{2^m-1} < 2$, esto es cuando el valor pertenece al primer nivel sin ser el máximo del nivel, puesto que si $i_{(m,n)}$ es el q-ésimo valor del primer nivel, el puede expresarse como,

$$i_{(m,n)} = 2^{m-q}[2^{q+1} - 1] - 1,$$

que igualmente no requiere acudir al siguiente paso, porque

$$j = 1 - \frac{1}{log_e 2}\left\{log_e\left\{\frac{2^{m+1} - [i_{(m,n)} + 1]}{2^m}\right\}\right\}.$$

Paso 2. Iteraciones para determinar el subíndice correspondiente a $i_{(m,n)}$.

Afirmar que $i_{(m,n)}$ pertenece al nivel r, es equivalente a afirmar que su subíndice j es tal que $D_{r-1}(m) < j \leq D_r(m)$, de manera que si $i_{(m,n)}$ es el q-ésimo elemento del nivel r, siendo $q = 1, 2, 3, ..., d_r(m)$, entonces $j = D_{r-1}(m) + q$. Para efectos de planteamiento del procedimiento iterativo de determinación del subíndice, el q-ésimo elemento del nivel r denotado por $l_{r,q}^{(n,m)}$ puede expresarse como,

$$l_{r,q}^{(n,m)} = l_q^{(r,m-1)} + 2^{m+r-1}$$

donde $l_q^{(r,m-1)}$ es el q-ésimo elemento del recorrido de la estadística $I_{(m-1,r)}$. Significa esto que $i_{(m,n)} - 2^{m+r-1}$ es el q-ésimo valor del recorrido de la estadística basada en tamaños de muestra $m-1$ y r, luego se estaría en una situación afín a la inicial del Paso 1, excepto que los tamaños de muestra se han modificado; el de la primera población se ha reducido en una unidad y el tamaño de la segunda a lo sumo es igual a n. El procedimiento regresa al paso uno, con $m-1, r$ y el valor $i_{(m-1,r)} = i_{(m,n)} - 2^{m+r-1}$, además con la certeza que $D_{r-1}(m) < j$.

3.3. Ejemplos

Ejemplo 1.

Asumiendo $m = 6, n = 4$, que la sucesión de rachas fue 1111101000, el valor particular de $I_{(6,4)}$ es,

$$\begin{aligned} i_{(6,4)} =& 1000 \\ =& 1 * 2^9 + 1 * 2^8 + 1 * 2^7 + 1 * 2^6 + 1 * 2^5 + \\ & 0 * 2^4 + 1 * 2^3 + 0 * 2^2 + 0 * 2^1 + 0 * 2^0. \end{aligned}$$

Los datos de entrada para esta primera iteración son: $m = 6, n = 4$, e $i_{(6,4)} = 1000$. El mínimo de la estadística es $2^6 - 1 = 63$. Como $\frac{1000}{63} < 2^4$, indica que 1000 pertenece al nivel $r = 4$ y no es su valor máximo; su subíndice es superior a 84, $\left(D_3(6) = \frac{9!}{6!3!} = 84\right)$. Como $i_{(6,4)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(5,4)} = 1000 - 2^{6+4-1} = 488$.

Los datos de entrada para esta segunda iteración son: $m = 5, n = 4$, e $i_{(5,4)} = 488$. El mínimo de la estadística es $2^5 - 1 = 31$. Como $\frac{488}{31} < 2^4$, indica que 488 pertenece al nivel $r = 4$ y no es su valor máximo; su subíndice es superior a 84+56=140, $\left(D_3(5) = \frac{8!}{5!3!} = 56\right)$. Como $i_{(5,4)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(4,4)} = 488 - 2^{5+4-1} = 232$.

Los datos de entrada para esta tercera iteración son: $m = 4, n = 4$, e $i_{(4,4)} = 232$. El mínimo de la estadística es $2^4 - 1 = 15$. Como $\frac{232}{15} < 2^4$, indica que 232 pertenece al nivel $r = 4$ y no es su valor máximo; su subíndice es superior a 140+35=175, $\left(D_3(4) = \frac{7!}{4!3!} = 35\right)$. Como $i_{(4,4)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(3,4)} = 232 - 2^{4+4-1} = 104$.

Los datos de entrada para esta cuarta iteración son: $m = 3, n = 4$, e $i_{(3,4)} = 104$.

El mínimo de la estadística es $2^3 - 1 = 7$. Como $\frac{104}{7} < 2^4$, indica que 104 pertenece al nivel $r = 4$ y no es su valor máximo; su subíndice es superior a 175+20=195, $\left(D_3(3) = \frac{6!}{3!3!} = 20\right)$. Como $i_{(3,4)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(2,4)} = 104 - 2^{3+4-1} = 40$.

Los datos de entrada para esta quinta iteración son: $m = 2, n = 4$, e $i_{(2,4)} = 40$.

El mínimo de la estadística es $2^2 - 1 = 3$. Como $\frac{40}{3} < 2^4$, indica que 40 pertenece al nivel $r = 4$ y no es su valor máximo: su subíndice es superior a 195+10=205, $\left(D_3(2) = \frac{5!}{2!3!} = 10\right)$. Como $i_{(2,4)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(1,4)} = 40 - 2^{2+4-1} = 8$.

Los datos de entrada para esta última iteración son: $m = 1, n = 4$, e $i_{(1,4)} = 8$.

El mínimo de la estadística es $2^1 - 1 = 1$. Como $\frac{8}{1} = 2^3$, indica que 8 pertenece al nivel $r = 3$ y **es su valor máximo**; su subíndice es finalmente $j = 205 + 4 = 209$, $\left(D_3(1) = \frac{4!}{1!3!} = 4\right)$.

Ejemplo 2.

Asumiendo $m = 5, n = 5$, que la sucesión de rachas fue 1000001111, el valor particular de $I_{(5,5)}$ es,

$$
\begin{aligned}
i_{(5,5)} =& 527 \\
=& 1 * 2^9 + 0 * 2^8 + 0 * 2^7 + 0 * 2^6 + 0 * 2^5 + \\
& 0 * 2^4 + 1 * 2^3 + 1 * 2^2 + 1 * 2^1 + 1 * 2^0.
\end{aligned}
$$

Los datos de entrada para esta primera iteración son: $m = 5, n = 5$, e $i_{(5,5)} = 527$.

El mínimo de la estadística es $2^5 - 1 = 31$. Como $\frac{527}{31} < 2^5$, indica que 527 pertenece al nivel $r = 5$ y no es su valor máximo; su subíndice es superior a 126, $\left(D_4(5) = \frac{9!}{4!5!} = 126\right)$. Como $i_{(5,5)}$ no pertenece al nivel 1, continúa en la iteración siguiente, con $i_{(4,5)} = 527 - 2^{5+5-1} = 15$.

Los datos de entrada para esta segunda y última iteración son: $m = 4$, $n = 5$, e $i_{(4,5)} = 15$.

El mínimo de la estadística es $2^4 - 1 = 15$. Como $\frac{15}{15} = 2^0$, indica que 15 pertenece al nivel $r = 0$ y **es su valor máximo**; su subíndice es por lo tanto $j = 126 + 1 = 127$, $(D_0(4) = 1)$.

Ejemplo 3.

Asumiendo $m = 6, n = 4$, que la sucesión de rachas fue 0001111101, el valor particular de $I_{(6,4)}$ es,

$$\begin{aligned} i_{(6,4)} &= 125 \\ &= 0 * 2^9 + 0 * 2^8 + 0 * 2^7 + 1 * 2^6 + 1 * 2^5 + \\ &\quad 1 * 2^4 + 1 * 2^3 + 1 * 2^2 + 0 * 2^1 + 1 * 2^0. \end{aligned}$$

Los datos de entrada para esta primera iteración son: $m = 6, n = 4$, $i_{(6,4)} = 125$.

El mínimo de la estadística es $2^6 - 1 = 63$. Como $\frac{125}{63} < 2^1$, indica que 125 pertenece al nivel $r = 1$ y no es su valor máximo; su subíndice es,

$$j = 1 - \frac{1}{log_e 2}\left\{log_e\left\{\frac{2^{6+1} - (125+1)}{2^6}\right\}\right\} = 6\,.$$

Ejemplo 4.

Una muestra de tamaño 5 de una primera población y otra muestra de tamaño igualmente 5 de una segunda población, fueron seleccionadas con el propósito de evaluar si las poblaciones difieren significativamente en cuanto a su localización.

Asumiendo el sistema a. de hipótesis, es decir,

$$H_o : \theta = 0$$

$$vs. \quad H_1 : \theta \neq 0,$$

los elementos del test específico son:

$\Omega(m,n) = \Omega(5,5) = 252$. Asumiendo $k = 6$, para tener un valor de $\alpha = \frac{12}{252}$ el cual se aproxima a 0.0476, $\Omega(m,n) - (k-1) = 252 - (5) = 247$, si la sucesión de rachas fue 1000001111, como se consideró en el Ejemplo 2, por lo tanto $i_{(5,5)} = 527$, que el mismo ejemplo concluye que su subíndice es $j = 127$. Estos argumentos no muestran evidencias estadísticas que permitan rechazar la hipótesis, a la luz de lo manifestado por el test γ_a. Adicionalmente, los llamados valores p, son fácilmente calculables; para este ejemplo particular $P[I_{(5,5)} \geq 527] = \frac{252-126}{252} = 0.5$

BIBLIOGRAFÍA

1. Gibbons, J. D. (1971), *Nonparametric Statistical Inference*, International Student Edition. McGraw-Hill Kogakusha, ltda, Tokyo.
2. Mayorga, J. H. (1993), *Un método de discriminación en dos grupos por medio de variables dicotómicas usando desarrollo binario*, Revista Colombiana de Estadística, 27, 26-38., Santafé de Bogotá (Colombia).