

DISTRIBUCIÓN DEL ÍNDICE DE ACUERDOS EN DIAGNÓSTICOS A , VIA *BOOTSTRAP*

Elizabeth Torres Rivas
Universidad de Los Andes
Instituto de Estadística Aplicada y Computación

Resumen. Los análisis de diagnósticos clínicos o epidemiológicos son a menudo imprecisos, presentan mucha variabilidad. Esto cuestiona la validez y precisión de los resultados en dichos estudios, pues a mayor variabilidad, mayor error y menor precisión.

Se propuso el índice de Acuerdo (A) para determinar el acuerdo interexaminador e intraexaminador, pues a mayor acuerdo menor variabilidad y viceversa (Torres y Dávila, 1994). Dado que la distribución de este índice está en estudio, se presenta aquí la estimación de su distribución. En casos como éste se puede aplicar tanto el método de Montecarlo como el *Bootstrap*. Se diseña un experimento donde se simulan, artificialmente los datos, vía Montecarlo, es decir, se generan m muestras (primarias) de igual tamaño. Luego, estas muestras primarias son remuestreadas por *Bootstrap*. En ambos casos se calcula el respectivo valor del índice A y su función de distribución muestral empírica. Al comparar las funciones de distribución, éstas resultan similares, aunque la obtenida por *Bootstrap* resulta más suave. Se determinan algunos intervalos "percentil" para ambos métodos y se observa que, a medida que aumentan las probabilidades de diagnóstico correcto, los intervalos se hacen más angostos, es decir, más precisos.

Introducción

El campo de la Estadística no Paramétrica ha incrementado su atractivo en los últimos años, con las nuevas herramientas

para análisis estadístico. Las cuales ofrecen sofisticadas alternativas a los modelos paramétricos tradicionales, para explotar grandes cantidades de datos univariantes o multivariantes, sin hacer supuestos distribucionales específicos. Entre dichas técnicas figuran los Métodos de Montecarlo y *Bootstrap*.

La variabilidad en los diagnósticos de los examinadores debe ser controlada cuando se realizan las encuestas epidemiológicas. Dicha variabilidad se cree se deba a: los juicios hechos por los examinadores en el diagnóstico de enfermedades que no son fácilmente detectables inicialmente o a la interpretación y aplicación de los criterios de diagnóstico, o factores psicológicos y físicos como fatiga o variaciones visuales y táctiles.

La variabilidad del examinador se puede definir como el fracaso del examinador para medir o identificar un fenómeno exactamente, lo cual produce un error. Las causas de esto, pueden deberse a: una falla del examinador, o a una técnica de medición defectuosa (medición incorrecta del test) o a una mala interpretación de los datos o de los criterios de diagnóstico.

Existen dos tipos de variación: **variación interexaminador**, se refiere a la variabilidad entre todos los examinadores cuando reportan el mismo material, y la **variación intraexaminador** que se relaciona a la variabilidad intrínseca de cada evaluador cuando evalúa más de una vez el mismo material. Aquí se estudia el primer tipo de variación.

Estos dos tipos de variación cuestionan la validez de los resultados, tanto en estudios epidemiológicos como clínicos, pues a mayor variabilidad, mayor error y menor precisión [Torres y Dávila, 1994]. Es decir, se desea que el grupo de examinadores realicen los diagnósticos con criterios uniformes y estándares. Para ello se realiza la calibración, la cual consiste en que el grupo de examinadores evalúe a cada uno de los pacientes o sujetos varias veces y se determine si es consistente, consigo

mismo (variación intraexaminador) y con sus colegas (variación interexaminador).

Los propósitos de la calibración son: evaluar la habilidad de los examinadores para realizar trabajos específicos consistentemente, asegurar una interpretación uniforme, para entender y aplicar los criterios de diagnóstico en las diferentes enfermedades y en las condiciones a ser observadas y registradas, y por último, minimizar las variaciones entre diferentes examinadores [Goulet y Flack, 1993].

Para la calibración se entrenan los examinadores con los mismos criterios de diagnóstico, se elige una muestra de pacientes a los que cada uno de los examinadores diagnostica.

En el artículo de Torres y Dávila (1994) se propuso el Índice de acuerdo A^1 como método para determinar el acuerdo interexaminador y intraexaminador, pero aún no se conoce la distribución del estadístico A , porque no se ha derivado analíticamente su distribución.

En el presente artículo se propone estudiar dicha distribución experimentalmente vía Montecarlo y *Bootstrap*, utilizando como base el Teorema de Glivenko-Cantelli.

El Índice A se calcula a partir de una tabla de doble entrada que contiene en la i -ésima fila y j -ésima columna los examinadores que realizan los diagnósticos (véase Tabla 1). En las celdas se localizan las coincidencias S_{ij} de todos los sujetos evaluados por cada par de examinadores (i,j) .

Tabla 1
Matriz de coincidencias interexaminador

i/j	1	2	3	...	l
1	S_{11}	S_{12}			S_{1l}

¹ La fórmula de A fue ideada por Gerard Defives, profesor de la Facultad de Ingeniería, ULA, Mérida en 1993, coordinador del grupo de investigación de Análisis de Datos.

2		S ₂₂	S ₂₃		S ₂₁
3			S ₃₃		S ₃₁
...			
1-1					S ₍₁₋₁₎₁
1					S ₁₁

Fuente: Propia.

Dicho Índice A se define como:

$$A = \frac{1}{\frac{l(l-1)}{2} S_{max}} \sum_{i=1} \sum_{j=2} S_{ij} \quad \text{para } i < j, i = 1, 2, \dots, l \quad [1]$$

el cual es el cociente de la suma de las coincidencias S_{ij} entre todos los pares de examinadores i y j en los diagnósticos, dividido entre el número de pares posibles de examinadores $(l(l-1)/2)$ y el número máximo S_{max} de acuerdo para un par de examinadores [Torres y Dávila, 1993]. Esta tabla tiene la forma de una matriz triangular, que contiene todos los posibles pares de combinaciones $(l(l-1)/2)$ de examinadores.

Para el cálculo se excluyen los elementos de la diagonal principal. Teóricamente, el índice A puede variar entre 0 y 1; donde 0 indica no acuerdo y 1, perfecto acuerdo, este es el caso de cuando todos los pares posibles de examinadores logran el máximo acuerdo [Torres y Dávila, 1993]. Un valor mayor que cero (pero menor o igual que uno), ya indica acuerdo.

1 Materiales y métodos

La metodología empleada para la generación artificial de los datos es la simulación a través del Método de Montecarlo y para el remuestreo se emplea el Método *Bootstrap*.

El Método Montecarlo fue introducido por Barnard [1963], para evaluar la importancia de un test estadístico, pues, se utiliza para contrastar la hipótesis nula de que los datos son una

muestra aleatoria de una población específica. Esto es logrado por la selección simulada de muestras de una población específica y comparando los valores del test estadístico de las muestras simuladas con el valor del test estadístico de la muestra real. Este método es de particular importancia en situaciones donde la distribución de la población es conocida, pero la distribución de la muestra del estadístico no ha sido derivada analíticamente [Noreen, 1989].

En el presente trabajo, se propone encontrar la distribución muestral del estadístico sin usar métodos analíticos.

El método *Bootstrap*, cuyo pionero fue Efron [1979], es de reciente desarrollo, puede ser utilizado cuando el objetivo es extraer conclusiones acerca de una población a partir del conocimiento de una muestra. En este método, las muestras artificiales son extraídas con reposición de la misma muestra. Dichas muestras *Bootstrap* pueden ser utilizadas de varias maneras: para estimar el nivel de significación de un test estadístico, estimar la función de distribución muestral empírica y la estimación de intervalos de confianza, que pueden ser modificados para estimar los niveles de significación. Aunque el método *Bootstrap* es muy fácil de usar y extremadamente flexible, se requiere precaución para su uso correcto. Sin embargo, esta metodología posee una gran ventaja sobre los métodos paramétricos convencionales; ellos pueden ser usados en situaciones en las cuales la distribución muestral paramétrica del estadístico es desconocida [Efron, 1993 y Noreen, 1989].

Para responder a la pregunta ¿Cómo se distribuye el Índice A ? se plantea examinar la función de distribución empírica, utilizando en el Teorema Glivenko-Cantelli [Lòeve, 1963] que dice:

Teorema Central de la Estadística. Si F es la función de distribución de una variable aleatoria X y F_n es la función de distribución empírica de X en n ensayos independientes e idénticos, entonces:

$$P[\sup_{-\infty < x < +\infty} |F_n(x) - F(x)| \rightarrow 0] = 1$$

En otras palabras, $F_n(x) \rightarrow F(X)$ uniformemente en x , con probabilidad 1.

La función de distribución muestral empírica, llamada F_n es un estimador de toda la población F , calculada a partir de una muestra aleatoria seleccionada de F . Esto es también conocido como principio *plug-in*, el cual es un método simple de estimación de parámetros a partir de muestras y el *Bootstrap* es una aplicación directa de dicho principio (Efron, 1993). En el remuestreo se usa una extensión de dicho teorema, demostrada por Del Canto (1992). Desde 1984 Del Canto ha venido desarrollando la fundamentación teórica del remuestreo (1984, 1985, 1995-a, 1995-b, Del Canto y Sinha [1994])

1.1 Función de distribución del índice A

Con base en el teorema anterior se define la función de distribución del índice A . Este índice es una variable aleatoria y se distribuye con la siguiente función de Distribución:

$$F_A(a) = P(A \leq a) \quad \forall a \in \mathfrak{R}$$
$$F_A(a) = \frac{1}{n} \sum I_{(-\infty, a]}(a) = \frac{1}{n}(n^* \leq a) \quad [2]$$

Donde, n^* es el número de observaciones de a_{ij}^* .

1.2 Función de distribución muestral empírica de A

Utilizando el Teorema de Glivenko-Cantelli se estima la función de distribución poblacional mediante la función de distribución muestral empírica, es decir:

$$F_n(a) = \hat{F}_A(a) \quad \forall a \in \mathfrak{R}$$
$$F_n(a) = \frac{1}{n} \sum I_{(-\infty, a]}(A_i) \quad [3]$$

Sean l examinadores que diagnostican una posible enfermedad, con cuatro categorías de respuestas (1, 2, 3,4), por ejemplo, a n pacientes. Una **muestra aleatoria** de tamaño n se representa por la matriz $X_{n \times l}$ de dimensión $n \times l$.

$$X_{n \times l} = [X_{kj}]$$

donde: X_{kj} es el diagnóstico dado por el j -ésimo examinador al k -ésimo individuo, para $k = 1, 2, \dots, n$ y $j = 1, 2, \dots, l$. Además

X_{kj} puede tomar los valores 1, 2,3 y 4 en este ejemplo. Con dos columnas cualesquiera de $X_{n \times l}$ se calcula el número S_{ij} de coincidencias del examinador i con el j . Con S_{ij} se obtiene el índice de acuerdo entre los examinadores según (1).

Esta muestra se replica m veces por el método de Montecarlo y para ellas sus estadísticos de orden son $a_{(1)}, \dots, a_{(m)}$ con las cuales se construye la función de distribución muestral Empírica y su respectivo gráfico a partir de:

$$F_m(a) = \frac{1}{m} \sum I_{(-\infty, a]}(a_i) = \frac{1}{m} (n^* \leq a) \quad [4]$$

Donde, n^* es el número de observaciones de a_{ij}^* .

Las funciones [3] y [4] parecen similares, pero no lo son, ya que [3] es una variable aleatoria, mientras que [4] es un valor específico, un caso particular que permite realizar el gráfico.

1.3 Función de distribución muestral empírica de a^* para las muestras *Bootstrap*

De cada muestra Montecarlo se obtienen B muestras *Bootstrap* y se calcula, para cada una de ellas, su respectivo Índice A^* y su función de distribución evaluada para los mismos valores de A en Montecarlo.

La función de distribución muestral empírica mancomunada² $F_{n,n}(a)$ se calcula así:

² La fórmula de esta función fue ideada por Dr. Surendra Sinha, Profesor Titular Jubilado del Instituto de Estadística Aplicada y Computación de FACES, ULA, Mérida.

$$F_{n,n}(a) = \frac{1}{m} \sum_{i=1}^m \frac{1}{B} \sum_{j=1}^B I_{(-\infty, A_{(i)}]}(a_{ij}^*) \quad \forall a \in \mathfrak{R}$$

$$F_{n,n}(a) = \frac{1}{m} \sum_{i=1}^m \frac{1}{B} \sum [n^* \leq a] \quad [5]$$

donde:

m = es el número de muestras Montecarlo

B = es el número de muestras *Bootstrap*

n^* = es el número de observaciones de a_{ij}^*

Además se puede simplificar [5] haciendo:

$$Z_{ij} = \frac{1}{B} \sum_{j=1}^B n^* \leq a \quad [6]$$

Donde, n^* es el número de observaciones a_{ij}^*

Luego la función [5] queda:

$$F_{n,n}(a) = \frac{1}{m} \sum_{i=1}^m Z_{ij} \quad [7]$$

Cada Z_{ij} se calcula como un promedio en las B muestras *Bootstrap* de los que son menores o iguales a $A_{(i)}$; éste es un valor obtenido en la muestra Montecarlo o elegido arbitrariamente.

La función [5] ó [7] se denomina función de distribución mancomunada, ya que, esta función de distribución se calcula para los valores del Índice de acuerdo (A) generados por el método Montecarlo, como puede observarse en la tabla 2.

Se generan m muestras Montecarlo, y para cada una de ellas, B muestras *Bootstrap*. Con estos valores ordenados, se observan cuántos valores de cada una de las réplicas *Bootstrap* son menores o iguales a un valor de A elegido.

Tabla 2
Función de distribución mancomunada

Orden	1	2	...	m	$F_{A,n}(A)$
$A_{(1)}$	Z_{11}	Z_{21}	...	Z_{m1}	$F_{A,n}(A_{(1)}) = \frac{1}{m} \sum_{i=1}^m Z_{i1}$
$A_{(2)}$	Z_{12}	Z_{22}	...	Z_{m2}	$F_{A,n}(A_{(2)}) = \frac{1}{m} \sum_{i=1}^m Z_{i2}$
...	
$A_{(m)}$	Z_{1m}	Z_{2m}	...	Z_{mm}	$F_{A,n}(A_{(m)}) = \frac{1}{m} \sum_{i=1}^m Z_{im}$

Fuente: Propia.

La primera columna de la Tabla 2 se refiere a los valores ordenados del índice A , obtenido por Montecarlo y la última es el promedio de las filas, esta es la Función de Distribución Mancomunada. Las columnas 1,...,m son los Z_{ij} para cada una de las muestras *Bootstrap* correspondiente a cada muestra Montecarlo.

1.4 Diseño del experimento

Para estudiar la función de distribución muestral empírica del índice de acuerdo (A) se diseña y simula por Montecarlo un experimento, donde se establecen las siguientes premisas:

- Las muestras provienen de una población X de diagnósticos que se distribuye $F(x)$.
- Existe una enfermedad que tiene, por ejemplo, cuatro categorías de respuesta mutuamente excluyente y colectivamente exhaustiva; existe una prevalencia de la enfermedad en la población, es decir, una probabilidad para cada categoría de respuestas. La suma de dichas probabilidades es igual a uno.
- Se utilizan, por ejemplo, $l = 6$ examinadores que van a evaluar a cada uno de los sujetos.
- Los examinadores utilizan los mismos criterios de diagnóstico, por lo tanto, se fija un único patrón de diagnóstico. Es decir, tienen las mismas probabilidades de diagnosticar correcta o incorrectamente a cada paciente o sujeto. El diagnóstico correcto tiene probabilidad p y el incorrecto $q = (1 - p) / 3$ (como son cuatro categorías de diagnóstico, el incorrecto se reparte en forma equitativa entre las tres restantes). Se fijan como probabilidades de diagnóstico correcto: desde 0.90 a 0.99, las cuales varían en 0.01, bajo el supuesto que los examinadores hayan sido entrenados previamente.
- Existe independencia entre los examinadores en la realización de sus diagnósticos.
- Lo anterior caracteriza la población y nos permite emplear el método de muestreo de Montecarlo.
- Se supone que las muestras Montecarlo son extraídas de la población y sirven de base para realizar el remuestreo por *Bootstrap*, de acuerdo al algoritmo *Bootstrap* [Efron, 1993].
- Por cada muestra Montecarlo extraída por simulación de la población, se seleccionan B muestras *Bootstrap*, de igual tamaño n , cuyas respectivas funciones de distribución se mancomunan, para comparar ambos métodos en igualdad de condiciones.

2 Resultados

Los datos generados artificialmente por Montecarlo, de acuerdo al experimento planteado para cada una de las probabilidades de diagnóstico correcto, son considerados como muestras primarias, a partir de las cuales son seleccionadas con reposición las muestras *Bootstrap*. Según esto, 100 muestras Montecarlo de tamaño 50 y por cada una de ellas, 50 muestras *Bootstrap* de igual tamaño, son seleccionadas. En ambos casos, el índice es calculado para cada una de las muestras.

Inicialmente se simularon réplicas Montecarlo y *Bootstrap* más grandes, pero se presentaron problemas de espacio en disco y en la capacidad de memoria del computador para el procesamiento de los datos completos.

En vista de esto, se decidió generar, para cada una de las m muestras Montecarlo, B muestras *Bootstrap* de igual tamaño y así, se tiene la misma cantidad de información para ambos métodos.

También, son comparados estos métodos, en igualdad de condiciones con relación a las funciones de distribución, éstas son calculadas para 30 puntos elegidos en forma arbitraria e igualmente espaciadas, como un criterio neutral, a fin de no favorecer a un método en particular. Sin embargo, el mínimo y máximo de dichos puntos se eligen de los valores mínimo y máximo observados en ambos conjuntos.

2.1 Comparación de las funciones de distribución

Desde el punto de vista gráfico, al comparar las funciones de distribución empíricas de Montecarlo y *Bootstrap*, se observa que Montecarlo aporta más información que *Bootstrap*, pues este último solo aporta información promedio.

En general, para cada una de las probabilidades de diagnóstico correcto, en las muestras *Bootstrap* se observan valores del índice menores que en Montecarlo. Por esta razón

al calcular la función de distribución se observan valores iguales a cero para Montecarlo, mientras para *Bootstrap* no. Además, se observa que la función de distribución Montecarlo alcanza rápidamente el valor 1.

Si dividimos los valores del índice de Acuerdo (A) en 3 etapas se observa lo siguiente: 1) para los primeros valores la función de distribución *Bootstrap* es mayor que la de Montecarlo; 2) para los valores intermedios ambas distribuciones coinciden y, 3) para valores finales la función de distribución *Bootstrap* es inferior que la de Montecarlo. Al comparar los gráficos de acuerdo a las probabilidades de diagnóstico correcto de los examinadores, se tiene que, a medida que disminuye la probabilidad aumenta la variabilidad en los diagnósticos, pues los saltos en las funciones de distribución son más pequeños. Y también se observa, que la diferencia en los valores de estas distribuciones disminuyen (véanse los gráficos 1, 2, 3 y 4).

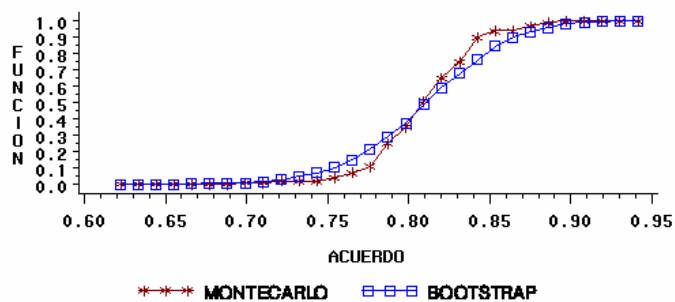


Gráfico 1. Comparación Montecarlo vs. Bootstrap
100 muestras Montecarlo y 50 muestras BOOTSTRAP
FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA (PROB. 0,90 DIAGNÓSTICO)

COMPARACION MONTECARLO VS. BOOTSTRAP

100 MUESTRAS MONTECARLO Y 50 MUESTRAS BOOTSTRAP
FUNCION DE DISTRIBUCION EMPIRICA (PROB. 0.91 DIAGNOSTICO)

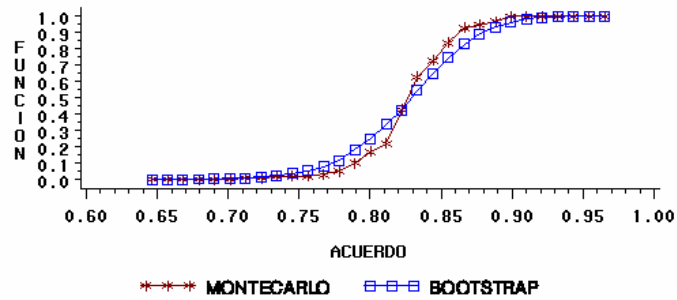


Gráfico 2. Comparación Montecarlo vs. Bootstrap
100 muestras Montecarlo y 50 muestras Bootstrap
Función de distribución empírica (prob. 0,91 diagnóstico)

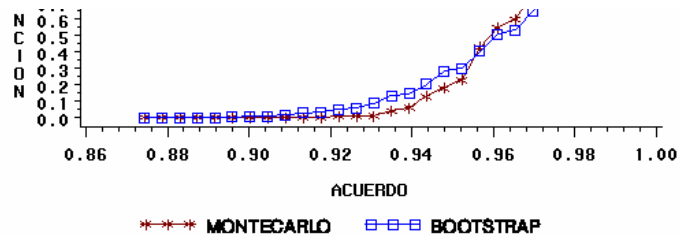
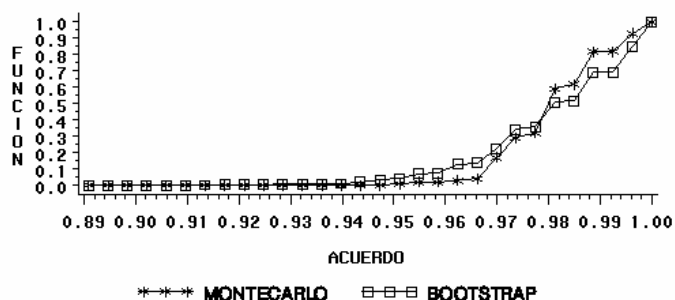


Gráfico 3. Comparación Montecarlo vs. Bootstrap
100 muestras Montecarlo y 50 muestras Bootstrap
Función de distribución empírica (prob. 0,98 diagnóstico)

COMPARACION MONTECARLO VS. BOOTSTRAP

100 MUESTRAS MONTECARLO Y 50 MUESTRAS BOOTSTRAP
 FUNCION DE DISTRIBUCION EMPIRICA (PROB. 0.99 DIAGNOSTICO)



**Gráfico 4. Comparación Montecarlo vs. Bootstrap
 100 muestras Montecarlo y 50 muestras Bootstrap
 Función de distribución empírica
 (prob. 0,98 DIAGNÓSTICO)**

Para verificar la similitud en el comportamiento de ambas funciones de distribución se aplica el análisis de correlación, para cada una de las probabilidades de diagnóstico correcto. Los resultados de dicho análisis indican que ambas funciones de distribución son similares, ya que, están fuertemente correlacionadas, con coeficientes de correlación altamente significativos, $\rho = 99\%$ ($p=0.001$), para todos los casos; esto se reafirma aún más, con las bajas y semejantes desviaciones estándares observadas.

Como se observa similar comportamiento en todos los casos analizados, esto indica que ni el tamaño de muestra ni la probabilidad de diagnóstico correcto influyen en la alta correlación observada entre dichas funciones. Es lógico que exista esa alta correlación, debido a que la función de distribución muestral empírica *Bootstrap* es obtenida a partir de las muestras Montecarlo.

2.1 Ajuste de las funciones según

distribución logística

Los datos obtenidos para las funciones de distribución Montecarlo y *Bootstrap* son ajustados a una Distribución Logística. La cual es definida en términos de su función de distribución:

$$F_x(x) = [1 + \exp\{-(x - \alpha)\}]^{-1} \text{ con } \beta > 0 \quad [8]$$

Las tablas 3 y 4 muestran el ajuste de las funciones de distribución para Montecarlo y *Bootstrap*, respectivamente, para todas las probabilidades. En ambos casos, se observa un buen ajuste por la Distribución Logística, debido a: 1) los cuadrados medios de los residuos son pequeños (0,10 y 0,08 respectivamente), 2) los estimadores de los parámetros Alfa y Beta tienden a 1 y 0, respectivamente, 3) los intervalos del 95% para los estimadores son precisos, tienen poca dispersión y 4) la correlación entre los estimadores de alfa y beta tiende a cero.

Tabla 3
100 muestras Montecarlo p= 0.90-0.99
ajuste distribución logística

Mínimos No Lineal Fuente	G.L.	Sumas de Cuadrados	Variable F ³ Dependiente C.M.
Regresión	2	62.991588459	31.495794229
Residual no corregido	298	29.832911541	0.100110441
Total	300	92.824500000	
(Total Corregido)	299	53.199363667	

Parámetro	Estimador	Error Estándar Asintótico Std.	Intervalo de Confianza Asintótico 95 %	
			Inferior	Superior
ALFA	0.9165879425	0.00511348373	0.90652469122	0.92665119384
BETA	0.0495432011	0.00585393376	0.03802275642	0.06106364582

Matriz de Correlaciones Asintóticas		
Corr	ALFA	BETA
ALFA	1	0.0597008999
BETA		0.05970089991

2.2 Intervalos percentil de 90, 95 y 99% de confianza

Con el uso del *Bootstrap* se pueden obtener intervalos muy precisos sin tener que hacer supuestos de normalidad. Los percentiles del histograma pueden utilizarse para definir los límites de los intervalos de confianza.

De acuerdo a la metodología propuesta por Efron (1993), el intervalo percentil se construye según el algoritmo siguiente: 1) se ordenan los valores del estadístico A , 2) se fija el nivel de significación y luego se calculan los percentiles para los límites inferior α y superior $1 - \alpha$.

³ F : Función de Distribución Montecarlo.

Tabla 4
bootstrap p=0.90-0.99
ajuste distribución logística

Mínimos No Lineal Fuente	G.L.	Sumas de Cuadrados	Variable FA1 ⁴ Dependiente C.M.
Regresión	2	60.839108090	30.419554045
Residual no corregido	298	25.072921990	0.084137322
Total	300	85.912030080	
(Total Corregido)	299	46.300557662	

Parámetro	Estimador	Error Estándar Asintótico Std.	Intervalo de Confianza Asintótico 95 %	
			Inferior	Superior
ALFA	0.9176446397	0.00495330664	0.90789661427	0.92739266517
BETA	0.0534540207	0.00582655277	0.04198746139	0.06492058011

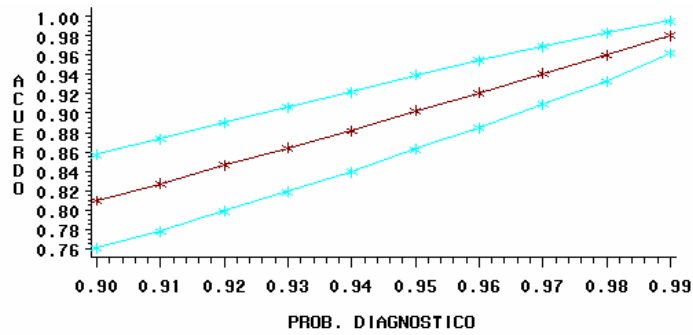
Matriz de Correlaciones Asintóticas		
Corr	ALFA	BETA
ALFA	1	0.0921081307
BETA	0.0921081307	1

Los gráficos para dichos intervalos se construyen comparando las probabilidades de diagnóstico con los acuerdos logrados. Por cada probabilidad de diagnóstico se calculan los percentiles de los límites inferior y superior para los niveles de confianza fijados en: 90, 95 y 99% respectivamente.

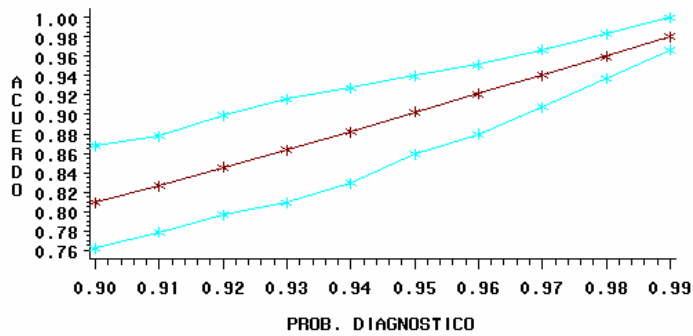
Al analizar los gráficos de ambos métodos, que se muestran a continuación, se observa que a medida que aumenta la probabilidad de diagnóstico correcto, el índice \hat{A} va aumentando, es decir, los acuerdos son mayores y además el intervalo se hace más angosto, produciendo intervalos más precisos, pues tienen menor dispersión. A medida que aumenta el nivel de confianza se hace más ancho el intervalo (véanse los gráficos 5, 6, 7, 8, 9 y 10).

⁴ FA 1 : Función de Distribución Bootstrap.

En el caso *Bootstrap*, los gráficos presentan los mismos patrones que en *Montecarlo*, pero las líneas de los intervalos son mucho más suaves.



**Gráfico 5. Intervalo percentil 90% (Bootstrap)
100 simulaciones y 50 Bootstrap**



**Gráfico 6. Intervalo percentil 90% (Montecarlo)
100 simulaciones**

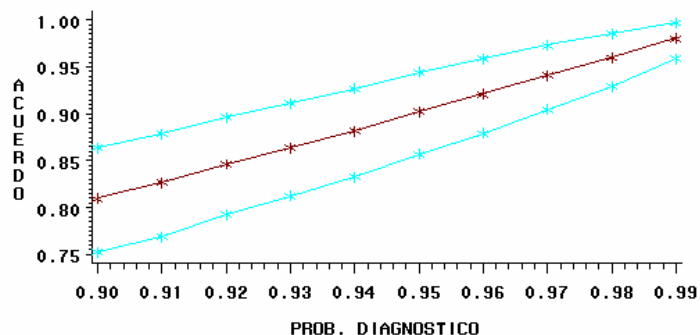


GRAFICO 4a.

Gráfico 7. Intervalo percentil 95% (Bootstrap)
100 simulaciones y 50 Bootstrap

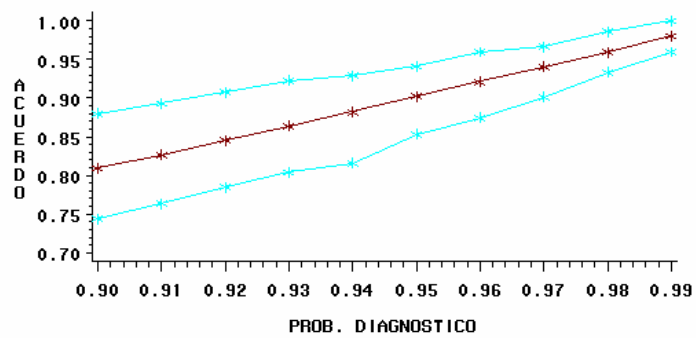


Gráfico 8. Intervalo percentil 95% (Montecarlo)
100 simulaciones

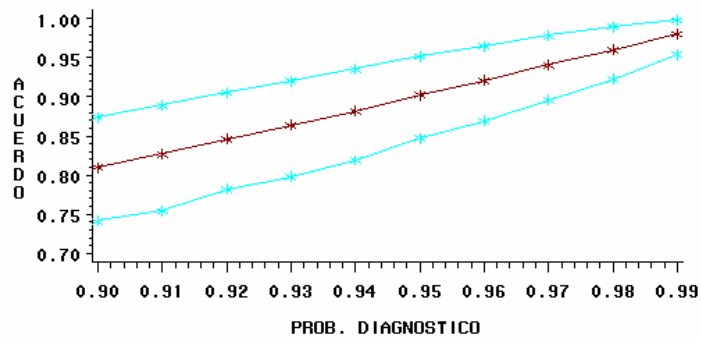


Gráfico 9. Intervalo percentil 99% (Bootstrap) 100 simulaciones y 50 Bootstrap

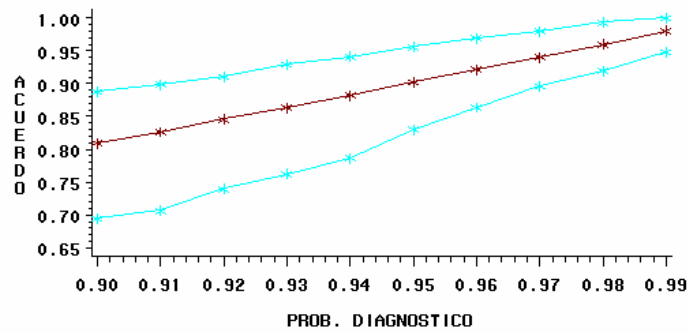


Gráfico 10. Intervalo percentil 99% (Montecarlo) 100 simulaciones

Torres, Elizabeth: Revista Economía No.13, 199., 205-228

3 Discusión

Se reporta el uso de un conjunto de técnicas de Diseño de Experimentos, Simulación, Bootstrap y Análisis de Datos, que denominamos Disianálisis⁵.

El experimento se diseñó considerando como tratamientos a las diferentes probabilidades de diagnóstico. Para generar los datos, es decir, los diagnósticos dados por los examinadores se utiliza simulación, específicamente, el Método de Montecarlo.

En las muestras Montecarlo se construye la función de distribución muestral empírica, mientras para las muestras *Bootstrap* se calcula la función de distribución empírica mancomunada, la cual es evaluada en los mismos puntos para Montecarlo. La función de distribución mancomunada cumple las propiedades de una función de distribución.

Dichas funciones son ajustadas a una Distribución Logística y al compararlas resultan similares, aunque la obtenida por *Bootstrap* resulta más suave.

Se determinan algunos intervalos "Percentil" del 90, 95 y 99% para ambos métodos y se observa que a medida que aumentan las probabilidades de diagnóstico correcto los intervalos se hacen más angostos, es decir, más precisos. Los gráficos para dichos intervalos, tanto de Montecarlo como de *Bootstrap*, nos muestran que, a medida que aumenta la

⁵ La conjunción de dichas técnicas, fue bautizada con el nombre Disianálisis por los profesores Renato del Canto y Surendra Sinha, en Septiembre de 1994, en una reunión de trabajo con ellos, el Prof. Carlos Domingo y yo. Por lo tanto, este trabajo es pionero en emplear el Disianálisis.

probabilidad de diagnóstico correcto, el índice A va aumentando, es decir, se logran mayores acuerdos y el intervalo se hace más angosto, produciendo intervalos más precisos. Por otra parte, a medida que aumenta el nivel de confianza los intervalos son más amplios.

El análisis de correlación evidencia que ambos métodos son igualmente eficaces para estudiar la distribución de este índice.

Inicialmente se simularon réplicas Montecarlo y *Bootstrap* más grandes, pero no se continuó por problemas de espacio en disco y en la capacidad de memoria del computador para la generación y procesamiento de los datos. Por tanto, es recomendable aumentar dichos tamaños, lo cual está planteado en la continuación de esta investigación.

Además, se plantea continuar la investigación utilizando otros métodos como: Jackknife y muestreo de muestras para ver si existe el mismo comportamiento en la función de distribución y si aparecen elementos de juicio que permitan decidir cuál método es mejor, o si se ratifica que son similares.

Bibliografía

- Barnard, G. A. (1963): **Journal of the Royal Statistical Society B.** 25, 294.
- Dávila, Balbina y Torres, Elizabeth (1993): “Calibración de los examinadores. Facultad de Odontología e Instituto de Estadística Aplicada y Computación”. FACES, ULA. ASOVAC. Noviembre. Mérida (Mimeografía).
- Del Canto J. Renato y Sihna, Surendra P. (1994): “Correlation between Sample Distribution Functions in Random Sampling and Random Resampling”. Instituto de Estadística Aplicada y Computación (IEAC), FACES, ULA. Ponencia presentada en: III Network of the Biometric Society-Caribbean, Central America, Colombia and Venezuela. June 27-July 1. Mimeografía. Caracas, Venezuela.
- Del Canto J., Renato (1984): “Introducción a la Teoría del Muestreo de Muestras (Caso Bernoulli)”. **Informe Técnico 840317-01**. IEAC, FACES, ULA. Mimeografía. Mérida.
- Del Canto J., Renato (1985): “Distribución Exacta del Vector Aleatorio $(F_r(x), F_{r,n}(x))'$ en el muestreo aleatorio de una muestra aleatoria”. Mimeografía. IEAC, FACES, ULA, Mérida.
- Del Canto J., Renato (1995-a): “Muestreo y re-muestreo en poblaciones finitas: Fundamentación”. IEAC, FACES, ULA, Mérida. 31 de Mayo. En homenaje al Dr. Surendra P. Sinha por su jubilación. Mimeografía.
- Del Canto J., Renato (1995-b): “Estadística Total: Presentación y Fundamentación”. IEAC, FACES, ULA. Abril. Mimeografía. Mérida.
- Del Canto, Renato (1992): “Fundamentación de la Inferencia Estadística basada en el muestreo de la muestra (Una

extensión del Teorema de Glivenko-Cantelli)". IEAC, FACES, ULA. Mimeografía. Mérida.

Dworkin, Samuel and Leresche, Linda (1992): "Research Diagnostic Criteria for Temporomandibular Disorders: Review, Criteria, Examinations and Specifications, Critique. J." **Craniomand Disorders**. 4: 318-25.

Efron, B. (1979) : "Bootstrap Methods : Another Look at the Jackknife". **Annals of Statistics**. 4 :1-26

Efron, B. and Tibshirani, R.(1993): "An introduction to the Bootstrap". **Monographs on Statistics and Applied Probability**. 57. Chapman & Hall. USA.

Goulet, G. and Flack, Clark (1993): "Reproductibility of Examiners Performance for Muscle and Joint Palpation int the Temporomandibular System following Training and Calibration. **Community Dent**". **Oral Epidemiol**. 21: 72-7.

Lòeve, Michel (1963): **Probability Theory**. Edit. D. Van Nostrand Company, Inc.

Noreen, Eric (1989): **Computer Intensive Methods for Testing Hypotheses. An Introduction**. John Wiley & Sons. New York.

Torres, Elizabeth and Dávila, Balbina (1994): "Variability Among Examiners in Oral Clinic Diagnoses". Instituto de Estadística Aplicada y Computación (IEAC), FACES, y Facultad de Odontología, ULA. Ponencia presentada en: III Network of the Biometric Society-Caribbean. Central América, Colombia and Venezuela. June 27-July 1. Mimeografía. Caracas, Venezuela.

Torres, Elizabeth y Dávila, Balbina (1993): "Variabilidad de los Examinadores". Instituto de Estadística Aplicada y Computación (IEAC). FACES y Facultad de Odontología.

Torres Elizabeth: Distribución del índice de acuerdos en...

ULA. ASOVAC'93. Noviembre. Mimeografía. Mérida,
Venezuela.