

**UNIVERSIDAD NACIONAL DE ROSARIO  
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA**

**CARRERA DE POSGRADO**

**DOCTORADO EN ESTADÍSTICA**

**Tema: Propiedades de índices de similitud y disimilitud para datos binarios y su aplicación en Ecología de comunidades**

**Autor: Patricia Susana Torres**

**Director: Dr. Sergio Camiz  
Directora: Dra. Marta B. Quaglino**

**2019**

Dedico esta tesis a mis componentes principales: a la memoria de mis abuelos Nenele y Francisco, a mis padres Rita y Hugo, a mis hermanos Claudia y Ricardo, a mis hijos Lara y Patricio, a mis hijos políticos Rodrigo y Agustina por querer a mis hijos tanto como yo y a mis sobrinos Paula y Francisco. Por último, pero no por eso menos importante, dedico esta tesis a la memoria de Paul, mi director de toda la vida, él quería que yo hiciera el Doctorado.

La suma de las partes multiplica la potencia.

Si quieres llegar rápido, camina solo,  
Si quieres llegar lejos, camina acompañado.

Proverbio africano

## AGRADECIMIENTOS

Quiero agradecer en primer lugar a mi director Dr. Sergio Camiz, por darme este tema, por su enorme ayuda, por su paciencia y por la amistad que logramos mientras trabajábamos juntos, en Rosario, en Mendoza, en Porto Alegre y en Roma. A mi co-directora Dra. Marta B. Quaglino por su gran ayuda en implementar los cálculos, las largas correcciones del manuscrito y la hermosa amistad que venimos disfrutando desde que era estudiante de Estadística.

Agradezco al Lic. César Mignone por su ayuda con las corridas de los programas de SAS. Agradezco al Dr. Valério Pillar quién simuló las coenoclinas que se utilizan en esta tesis.

Muchas gracias a mis compañeros Rodrigo Freire y Eugenio Alvarez Arnesi por la ayuda con las figuras en el programa R y a Graciela Klekailo por ayudarme con las tablas y la edición de la Tesis.

Agradezco a mi compañero Ignacio Barberis por su corrección de la Introducción. Agradezco a todos mis compañeros de trabajo por su enorme apoyo, a los más viejos, aquellos que ya se jubilaron pero compartieron gran parte de su vida conmigo, incluso los datos de vegetación que se utilizan en esta tesis. Gracias Nelín Carnevale, Silvia Boccanelli, Eduardo Franceschi y Eduardo Pire.

A mis nuevos compañeros, gracias José Vesprini, Mariana Mazzei, Jorgelina Assmus, Angeles Tessore, Caterina Barison y Leandro Sosa.

Agradezco a mis padres el cariño y el apoyo que siempre me dieron y la ayuda incondicional de mi madre Rita. A mis hijos Lara y Patricio por el apoyo y la paciencia mientras hacía cálculos y más cálculos. Por último agradezco la enorme compañía de Cayetana y Gringo siempre fieles y a mi lado mientras escribía esta tesis.

Como siempre sucede en estos casos, hay muchas otras personas que con su pequeña o gran ayuda hicieron posible esta tesis, para todas ellas va también mi auténtico y generoso agradecimiento.

## RESUMEN

Se estudian en esta tesis las propiedades de los índices de similitud basados en datos de presencia/ausencia, con el objetivo de identificar los “mejores”, basado en algún criterio teórico. Tales índices han sido ampliamente utilizados en ecología de comunidades para clasificar tanto datos de especies como caracteres ecológicos: índices idénticos o similares fueron aplicados subsecuentemente en otras disciplinas científicas. Como resultado del esfuerzo por cuantificar la asociación y/o similitud en varios campos de la Ciencia, han aparecido una gran cantidad de medidas. Si bien hay más de setenta índices para medir la similitud entre objetos, poco se sabe sobre las propiedades de los mismos: la revisión más completa de los coeficientes de similitud es debida a Hubálek (1982), limitada a 43. Siguiendo a este autor se estudian en este trabajo quince propiedades, en particular cinco son consideradas obligatorias: maximal y minimal, existencia de valores mínimos y máximos, simetría, discriminación entre asociación positiva o negativa y monotonidad respecto a la estadística  $\chi^2$ . Las otras propiedades que se estudian son: existencia de límites convencionales, asociación completa o absoluta, asociación nula cuando el valor de presencias conjuntas es cero, linealidad, insensibilidad a la inexistencia de presencias conjuntas, distribuciones en el muestreo, insesgamiento, métrica, indeterminancia y Euclideanidad. Se utilizaron diferentes métodos para comprobar las propiedades de los índices: algunas pueden verificarse sencillamente observando el cumplimiento de las condiciones en forma aritmética, para otras debe recurrirse a un procedimiento de simulación de datos binarios con distintos escenarios; para otras se utilizan los mismos escenarios definidos por otros autores que estudiaron alguna propiedad en un conjunto de índices. Se encuentran doce índices (A) que son los que más propiedades cumplen: Anónimo\_1, Anónimo\_2, Anónimo\_3, Braun-Blanquet, Hamann, Jaccard, Lamont & Grant, Maxwell & Pilliner, Rogot & Golberg, Scott, Sokal & Michener y Sorensen. Estos índices se presentan como los mejores para un uso general. Utilizando una medida de disimilitud (D) derivada de estos índices de similitud:  $D = \sqrt{1 - A}$ , se compara el comportamiento de la misma con tres

matrices de datos: una transecta de vegetación, y dos conjuntos de datos simulados: una coenoclina y un coenoplano, respondiendo a uno y dos gradientes ambientales hipotéticos respectivamente. No sorprende encontrar que los índices de Jaccard y Sorensen son adecuados para analizar datos de comunidades vegetales, sin embargo se descubren otros índices que también dan resultados satisfactorios: Braun-Blanquet, Lamont & Grant y Anónimo\_1.

**Palabras clave:** índices, similitud, propiedades, métrica, euclideanidad, multivariado, vegetación.

## ÍNDICE

RESUMEN	4
CAPÍTULO 1: INTRODUCCIÓN	8
OBJETIVO	19
Objetivos específicos	19
CAPÍTULO 2: METODOLOGÍA	21
2. 1. Notación	21
2. 2. Definiciones	23
2. 3. Propiedades de los índices de similitud y/o disimilitud	35
2. 4. Propiedades de las matrices de similitud y de disimilitud	45
2. 5. Diseño del estudio comparativo	48
CAPÍTULO 3: RESULTADOS Y DISCUSIÓN	54
CAPÍTULO 4: SELECCIÓN DE LOS MEJORES ÍNDICES A PARTIR DE SUS PROPIEDADES	105
CAPÍTULO 5: EFICIENCIA DE LOS MEJORES ÍNDICES PARA EL ANÁLISIS DE DATOS	122
5. 1. Descripción de las bases de datos	123
5. 2. Análisis multivariado sobre las bases de datos reales y simuladas	126
COMENTARIOS FINALES	136
BIBLIOGRAFÍA	137
ANEXO 1	153

ANEXO 2	171
ANEXO 3	189
ANEXO 4	219
ANEXO 5	239
ANEXO 6	255
ANEXO 7	259
ANEXO 8: PROGRAMAS DE COMPUTACIÓN	299

# CAPÍTULO 1

## INTRODUCCIÓN

El estudio de la vegetación a nivel de la comunidad plantea problemas considerables, ya que el conjunto de datos puede ser de gran dimensión y complejo. Un primer paso para su análisis suele ser la búsqueda de esquemas en la estructura y la composición de la comunidad. La necesidad de encontrar procedimientos para describir y comparar las comunidades ha dominado el desarrollo de la Ecología de las comunidades. En esencia, se trata de la búsqueda de modos simples de describir sistemas complejos (Begon *et al.* 1990).

La disciplina que se ocupa de la determinación y caracterización de comunidades vegetales, basado en sus atributos florísticos, es la Fitosociología. El método clásico fue desarrollado por el botánico suizo J. Braun-Blanquet (1928, 1932, 1979) y basa el análisis en la fidelidad de las especies, esto es, el grado de restricción de las especies a una comunidad. Una especie muy fiel es aquella que crece en una y sólo una comunidad, en cambio las especies poco fieles aparecen indistintamente en más de una o varias comunidades. El método de Braun-Blanquet significó un avance muy importante para la Fitosociología y permitió estudiar gran parte de la vegetación de Europa.

En Ecología de comunidades vegetales, el objeto de estudio suele ser un conjunto de datos compuesto por  $r$  unidades muestrales y  $n$  variables o especies. Esta matriz de datos en general es compleja, ya que representa a muchas variables, tiene gran cantidad de ceros y es de gran dimensión.

El tratamiento clásico de toda la información contenida en la matriz de datos, propuesto inicialmente por Braun-Blanquet (1932), consiste en reordenar las filas y las columnas de la matriz agrupándolas por su parecido hasta llegar a una tabla final ordenada. En la actualidad se utilizan distintos métodos de Análisis Multivariado, como recomienda Anderson (2001a) siguiendo el esquema de Clarke & Green (1988) y Clarke (1993) para el análisis de datos en Ecología.

Otros autores, como Legendre & Legendre (2012) sostienen que los métodos multivariados, diseñados para analizar conjuntos de datos complejos, teniendo en cuenta la estructura de correlación de los datos ecológicos, permiten evidenciar la estructura subyacente de los mismos. Peck (2010) afirma que una comunidad de plantas o animales está caracterizada por un número de especies cuya presencia o abundancia local depende de sus interacciones como así también de un sinnúmero de condiciones ambientales. Por esta razón, los ecologistas de comunidades necesitan herramientas que tengan en cuenta las múltiples respuestas, que a su vez no son independientes entre sí. Esta falta de independencia es el fundamento del análisis multivariado. Wildi (2010) habla de la naturaleza multivariada de los datos ecológicos, consistentes en varios atributos o unidades muestrales, lo cual hace necesario su tratamiento simultáneo.

Como declara Anderson (2001a), primero se debe decidir si es necesario una transformación y/o estandarización de los datos, en segundo lugar se debe elegir una medida de similitud o disimilitud para ser usada como la base del análisis y como tercer paso realizar un Ordenamiento y/o Clasificación que permita visualizar las relaciones entre los objetos de estudio basados en las especies. En ocasiones el interés de la investigación puede requerir la aplicación de un test multivariado con aleatorización (Pillar & Orlóci 1996) para comparar grupos definidos "a priori". Esto permite obtener una declaración probabilística rigurosa sobre los efectos multivariados de los grupos naturales.

En esta Tesis la atención se centra en un paso indispensable del análisis estadístico que se aplica cuando el objetivo es reducir y describir el conjunto de datos a través de técnicas de ordenamiento y/o clasificación, la selección de una medida de similitud o disimilitud adecuada al tipo de datos disponibles.

No serán abordados en este trabajo las posibles transformaciones y/o estandarizaciones de los datos, ni el desarrollo del test multivariado con aleatorización, a veces conocido como MANOVA no paramétrico (Edgington 1987, Efron & Tibshirani 1994, Manly 1998, Anderson 2001a y b, Torres et al. 2010).

Los métodos de clasificación, introducidos separadamente por Florek et al. (1951) y Sneath (1957b), tienen como objetivo definir clases o grupos discretos de los objetos o de las variables. La clasificación forma grupos o clases que son

homogéneas dentro y heterogéneas entre sí. Se pueden clasificar tanto los objetos como las variables o realizarlo simultáneamente (Mc Cune & Mefford 2016), lo cual arroja como resultado la tabla ordenada de vegetación (Braun-Blanquet 1979).

El término ordenamiento fue originariamente introducido en Ecología Vegetal por Goodall (1954) y definido como “la disposición de los objetos en un orden uni o multidimensional” y cuyo objetivo principal es representar la estructura de los datos en un espacio reducido. La finalidad del ordenamiento es representar los datos en una forma geométrica: los objetos y las variables son representados como puntos en un espacio reducido, de tal forma que las relaciones geométricas entre ellos se correspondan con las características del conjunto multidimensional. De tal modo se obtiene una descripción de las relaciones existentes entre los objetos o entre las variables de una forma accesible al ojo humano, que permite dar una interpretación del fenómeno, aspecto que en la mayor parte de los casos sería imposible o muy difícil de obtener directamente de los datos iniciales. Para que la configuración obtenida del análisis sea accesible al ojo humano, es necesario que la dimensión del espacio sea pequeña, de modo de permitir una representación gráfica, por ejemplo, en un plano.

Si bien las dos técnicas de ordenamiento y clasificación difieren entre sí con respecto al resultado que arrojan, tienen en común la elección de una medida de similitud o disimilitud para cuantificar la asociación entre objetos o entre variables. Existen muchas medidas de asociación, que varían según sea la naturaleza de los datos.

De acuerdo a como se haya realizado la colección de los datos, se puede disponer simplemente del registro de la presencia o ausencia de las especies o además de alguna medida de su importancia. Estas pueden ser la cobertura, la abundancia, la abundancia/cobertura combinada de la escala clásica de Braun-Blanquet (1932), frecuencia, biomasa, área basal, etc. La presencia/ausencia es una medida muy útil en estudios a gran escala (Mc Cune & Grace 2002). En estudios donde la heterogeneidad de las unidades experimentales es grande, la mayoría de la información de los datos será reportada solo por la presencia/ausencia de las especies. Sin embargo, la presencia/ausencia no es útil en detectar diferencias más sutiles en áreas más homogéneas. Por ejemplo

en una comparación de bosques de crecimiento lento versus bosques manipulados por el hombre para un crecimiento más rápido en Montana (USA), el líquen *Alectoria sarmentosa* estuvo presente en todas las unidades experimentales, pero fue consistentemente mucho más abundante en los bosques de crecimiento lento (Lesica & Steele 1997).

Las distintas medidas de asociación pueden evaluar diferentemente las relaciones entre objetos o entre variables y la elección de un coeficiente afecta los resultados del Análisis Multivariado posterior (Moore & Russell 1967). De ahí la importancia de elegir una medida de asociación apropiada, para lo cual Legendre & Legendre (1998) sugieren dos consideraciones importantes que deberían ser tenidas en cuenta. En primer lugar, la naturaleza del estudio, que es lo que determina el tipo de estructura a ser evidenciada mediante la matriz de asociación y por lo tanto el tipo de medida de asociación a utilizar y adicionalmente, las distintas restricciones matemáticas que satisfacen las medidas disponibles, ya que los métodos a los cuales será sometida la matriz de asociación con frecuencia requieren medidas de asociación con propiedades matemáticas específicas.

Otro tema a considerar es el aspecto computacional, que hoy ha perdido relevancia. En general los investigadores debían limitarse a utilizar los índices que estaban incorporados en los programas, sin embargo, hoy el enfoque para elegir un indicador debería priorizar sus propiedades. Un criterio para esta elección debería ser que el índice sea sencillo, refleje aquello que interesa al investigador y que tenga buenas propiedades.

Se han propuesto en la literatura una gran cantidad de coeficientes de asociación (Kenkel y Booth 1987, Hubálek 1982, Legendre & Legendre 1998), que según Giansanti (1990) se pueden clasificar en cuatro grandes grupos:

1.- Coeficientes de similitud y/o disimilitud (e.g. Sokal & Michener, Roger & Tanimoto, Sokal & Sneath, Hamann, Yule, phi de Pearson, Jaccard, Sorensen, Russell & Rao, Kulczynski, Ochiai, Gower, Bray-Curtis, etc.) Constituyen el grupo más grande de coeficientes (Legendre & Legendre 1998) y se utilizan para medir la asociación entre objetos o entre variables.

2.- Medidas de distancia (e.g. euclídea, cuerda, Mahalanobis) Constituyen una serie de coeficientes que tienen la característica común de cumplir con los axiomas de distancia. Asumen el valor máximo para dos objetos que resultan enteramente diferentes y el valor 0 para dos objetos que tienen los mismos caracteres. Son usados principalmente para medir asociación entre objetos comparando caracteres de tipo cuantitativo, aunque también se pueden utilizar para datos dicotómicos.

3.- Coeficientes de correlación (e.g. Pearson, Spearman, Kendall) Son utilizados principalmente para describir las relaciones entre los caracteres midiendo el grado de proporcionalidad y de dependencia entre las variables. Tales coeficientes pueden ser vistos como casos particulares de productos escalares, ya sea en el caso de variables cuantitativas, cualitativas, ordinales o dicotómicas.

4.- Coeficientes probabilísticos (e.g. Lance & Williams, Goodall) Han sido desarrollados más recientemente, e incluyen información de tipo estadística que mide el grado de homogeneidad del sistema particionando o sub-particionando conjuntos de objetos.

La similitud y la disimilitud son dos caras de la misma moneda y en la literatura se ha encontrado que el primer término se aplica a ambos (Rodríguez-Salazar et al. 2001). Un valor alto de un coeficiente de similitud indica que hay mayor asociación o parecido entre los dos objetos o las dos variables que se están comparando. Por el contrario, un valor bajo indica que los mismos son disímiles. La disimilitud es el complemento de la similitud, por eso, su interpretación es a la inversa.

Muchos de estos coeficientes fueron desarrollados en un principio para datos binarios y luego se extendieron a variables multiestados. También se encuentran en la literatura algunos índices para datos binarios que son casos especiales de fórmulas más generales para varios dominios de análisis (Warrens 2008). Por ejemplo, el índice de Sorensen (1948) también propuesto por Gleason (1920), Dice (1945) y Nei & Li (1979) para datos binarios, parece haber sido popularizado por Bray (1956) y Bray & Curtis (1957) y es un caso especial del coeficiente de Czekanowski (1932), de la medida de Odum (1950) y del coeficiente de Williams, Lambert & Lance (1966). El coeficiente de *simple matching* de Sokal & Michener (1958) puede ser obtenido por ejemplo como un caso especial de un coeficiente

más general atribuido a Gower (1971) o Cox & Cox (2000) o como un caso especial de una medida de Cain & Harrison (1958).

En este trabajo se consideran solamente coeficientes para datos binarios, los cuales son encontrados en muchas disciplinas (Warrens 2008). En Ecología de comunidades vegetales se ha demostrado la ventaja de trabajar con datos binarios (Camiz et al. 2017) en lugar de la clásica escala combinada de abundancia/cobertura de Braun-Blanquet (1979). Además varios tipos de datos cuantitativos pueden ser dicotomizados y tratados como binarios, en particular datos cuantitativos con ruido (Warrens 2008). Los coeficientes de asociación o similitud basados en datos binarios fueron utilizados por primera vez en Ecología para agrupar comunidades bióticas (Jaccard 1901) y especies ecológicamente relacionadas (Forbes 1907). Índices idénticos o similares fueron aplicados subsecuentemente en psicología, etología (Driver & Kroeber 1932), taxonomía (Sneath 1957a) y otras disciplinas biológicas. En las disciplinas que utilizan clasificación numérica para el análisis de comunidades o de especies (Cattell 1952, Sneath & Sokal 1973), se utilizan índices de similitud y de asociación para medir la asociación entre entidades y el comportamiento de dos descriptores respectivamente.

Cuando se mide la similitud entre dos objetos o variables, la atención está puesta principalmente en lo que tienen en común, es decir, en las presencias conjuntas. Sin embargo, no queda muy claro si lo opuesto, las ausencias conjuntas aportan a la similitud y es ésta la principal distinción que se hace entre los índices de similitud, si tienen en cuenta a esta última o no. Otra distinción que se hace entre los coeficientes de asociación es llamarlos coeficientes o índices de similitud propiamente dichos si varían entre  $[0,1]$ , mientras que a los que varían entre  $[-1,1]$  se los reconoce como coeficientes o índices de asociación (Clifford & Stephenson 1975). En este trabajo no se hace distinción entre ambos y se los nombra indistintamente.

Como resultado del esfuerzo por cuantificar la asociación o similitud en varios campos de la Biología, han aparecido una gran cantidad de medidas. La Taxonomía es la disciplina que mayormente ha basado sus propios estudios sobre la aplicación de tales coeficientes en métodos de análisis de datos. La sistemática del reino animal y del reino vegetal son sin duda las clasificaciones más grandes que ha hecho el hombre. El desarrollo de la Taxonomía numérica

a partir de la mitad de 1950 llevó a un rápido incremento en el uso y la construcción de coeficientes de asociación (Goodman & Kruskal 1959). Estos, sin embargo, han sido frecuentemente definidos sobre bases empíricas y sin una adecuada justificación teórica. Además, la insuficiencia de comunicación entre las distintas disciplinas ha llevado a numerosas duplicaciones de los mismos métodos en diferentes estudios (Giansanti 1990).

Si bien hay más de setenta índices para medir la similitud (o disimilitud) entre individuos, objetos o unidades experimentales, algunos de los cuales también se pueden emplear para comparar variables, poco se sabe sobre las propiedades de los mismos. Solamente se han realizado unos pocos estudios comparativos de la gran cantidad de coeficientes que se encuentran en la literatura (Simpson 1960, Sokal & Sneath 1963, T'Mannetje 1967, Goodall 1973, Sneath & Sokal 1973, Baroni-Urbani & Buser 1976, Austin & Colwell 1977, Janowitz 1980). Varios trabajos están dirigidos a analizar y confrontar estos coeficientes: Goodman y Kruskal (1954, 1959, 1963), Dagnelie (1960), Sokal y Sneath (1963), Cole (1949) y Cheetham y Hazle (1969), como así también para guiar al investigador en la elección de la medida de asociación adecuada al estudio de los fenómenos. Williams & Dale (1965) realizaron un análisis muy importante desde el punto de vista matemático. Gower y Legendre (1986) y Fichet y Le Calvé (1984) han presentado algunas propiedades de los coeficientes de similitud y en particular sus capacidades para producir matrices métricas y que cumplen con las propiedades de distancia euclídeana.

La revisión más completa de los coeficientes de similitud es debida a Hubálek (1982) quien publicó un trabajo presentando cuarenta y tres coeficientes de asociación (similitud y/o disimilitud), probablemente la colección más completa. Este autor señaló que algunos coeficientes son sinónimos de otros, algunos son múltiplos lineales o cuadráticos de otros, su logaritmo o su función trigonométrica. También algunos coeficientes son meras transformaciones desde un rango de valores a otro. El propósito de su trabajo fue comparar tanto en forma teórica como empírica una cantidad sustancial de coeficientes de similitud o asociación y seleccionar un grupo de ellos que podrían considerarse los más útiles y teóricamente los menos objetables. Para ello, definió cinco propiedades fundamentales que, a su parecer, debían cumplir los índices de

similitud para ser considerados admisibles: maximal y minimal, existencia de valores mínimos y máximos, simetría, discriminación entre asociación positiva o negativa y monotonía respecto a la estadística  $\chi^2$ . También definió siete propiedades más que, según su criterio, no son obligatorias, pero podrían ayudar en la selección de los mejores índices de asociación. Dichas propiedades son las siguientes: existencia de límites convencionales, asociación completa o absoluta, asociación nula cuando el valor de presencias conjuntas es cero, linealidad, insensibilidad a la inexistencia de presencias conjuntas, distribuciones en el muestreo e insesgamiento.

Su trabajo consistió en probar estas propiedades, en el conjunto de índices definidos, de los cuales solamente veinte resultaron admisibles y a los cuales clasificó con el método del promedio (Sokal & Michener 1958) y el método del *half-linkage* (Hubálek 1978). Los resultados de ambas clasificaciones son similares y se destacan cinco grupos de coeficientes. Los coeficientes de un mismo grupo produjeron resultados muy similares sobre la base de la asociación interespecífica de los datos estudiados y podrían ser considerados conjuntamente funciones monótonas. Por otro lado, los coeficientes desde diferentes grupos pueden producir distintas clasificaciones sobre los mismos datos, en efecto, esto es demostrado por Hubálek (1982) en los dendrogramas construidos con el método de ligamiento promedio con coeficientes representativos de cada uno de los grupos.

Una de las propiedades mencionada por Hubálek, pero no estudiada por él, se refiere a las distribuciones de los índices sobre una muestra de permutación grande. Solamente menciona en su trabajo una referencia al trabajo de Baroni-Urbani & Buser (1976) quienes encontraron que el índice desarrollado por ellos presentaba una distribución aproximadamente simétrica, desde una muestra de permutación grande de datos de presencia/ausencia, mientras que los índices desarrollados por Yule no estaban bien distribuidos, mostrando distribuciones asimétricas.

Otra propiedad definida en su trabajo pero tampoco estudiada es la propiedad de insesgamiento. En su trabajo menciona los resultados de Goodall (1973) quien observó que los índices de Jaccard y Yule son en general coeficientes insesgados, aún con muestras pequeñas, Sorensen y Sokal & Michener presentan algún sesgo, mientras que el índice de Cole mostró discrepancias

significativas entre los valores muestrales y los poblacionales, aún con muestras grandes.

Hubálek (1982) termina su trabajo, afirmando que otro argumento válido para la elección de un índice, aunque tampoco él lo estudia, podría ser la propiedad métrica de la disimilitud que resulta del complemento de la similitud. Esta propiedad se refiere a la posibilidad de representar puntos en un espacio métrico. Otra propiedad ligada a ésta y que puede ser estudiada en las matrices de disimilitudes es la euclideanidad, que permite la representación en un espacio euclideo, un espacio cuya métrica resulta de un producto escalar definido positivo. Esta condición permite representar correctamente un conjunto de objetos en un espacio gráfico a partir de una matriz de similitud (o disimilitud) entre los mismos y la elección de la técnica multivariada apropiada para lograr esta representación gráfica de ordenamiento depende del cumplimiento de esta propiedad. Gower & Legendre (1986) y Cuadras (1989) estudiaron las propiedades métricas y euclidianas de catorce índices de disimilitud derivados de similitudes conocidas. Legendre & Legendre (1998) estudiaron veintiséis índices, algunos de los cuales coinciden con los anteriores y son los más utilizados.

Entre otros autores que estudiaron algunas de las propiedades definidas por Hubálek, se menciona a Giansanti (1990). Esta autora trabajó con veintitrés índices, coincidentes en su mayoría con los veinte que habían resultado admisibles para Hubálek y estudió varias propiedades: la simetría, la linealidad, los valores máximos y mínimos y también las propiedades métricas y euclidianas. La mayoría de sus resultados coinciden con lo encontrado por Hubálek aunque con algunas discrepancias. Sin embargo, aún quedan numerosos índices frecuentemente utilizados en el área de Ecología para los cuales son desconocidas sus propiedades y no se desprende del trabajo de Giansanti ninguna comparación completa entre índices para seleccionar aquellos con las mejores propiedades. De aquí el interés de continuar con el estudio de las propiedades de los distintos índices de similitud/disimilitud disponibles en la literatura.

Un trabajo más reciente es debido a Warrens (2008), que estudió un problema de los coeficientes de similitud para datos binarios al que no se le había prestado mucha atención, y es que la mayoría de los índices son definidos como fracciones, por lo tanto el denominador puede hacerse cero, en cuyo caso el valor del coeficiente es indeterminado. Este autor presenta cincuenta y seis índices de similitud y estudia la indeterminancia para veintiocho de ellos, calculando el número de veces que se hacen indeterminados de acuerdo a un tamaño de vector fijo. Para estos casos, presenta nuevas definiciones siguiendo a Batagelj & Bren (1995), que proponen eliminar las indeterminancias definiendo valores apropiados en los casos críticos. El índice de Jaccard, por ejemplo, a veces puede ser indeterminado en una situación en la cual se comparan dos unidades experimentales que consisten enteramente de ceros, pero en Ecología, es improbable tener una tabla de datos (unidades experimentales x especies) que tenga unidades experimentales sin especies. Adicionalmente, el problema se puede resolver excluyendo los vectores de ceros de los datos. En su trabajo Warrens también estudia algunas propiedades de los índices de similitud para datos binarios, pero trabajando con familias de índices, en lugar de trabajar con los índices en forma individual. Define dos familias de índices, tal como fueron definidas también por Gower y Legendre (1986), una de las cuales no incluye las ausencias conjuntas y la otra sí. Además, define dos nuevas familias de índices desde las dos anteriores como combinaciones lineales y una quinta familia expresando nuevos coeficientes como medias aritméticas, geométricas y armónicas de ciertos coeficientes. Warrens (2008) considera que hay algunas ventajas en estudiar las familias de índices en lugar de hacerlo individualmente, dado que, desde la formulación de la familia, se puede deducir de qué manera están relacionados los diferentes miembros y además, las propiedades de los mismos, al igual que los límites, son fácilmente investigados. Otra ventaja es que se pueden obtener resultados más generales. En este trabajo se prefiere estudiar cada índice por separado y presentar los resultados individuales, considerando que se obtiene información individual para cada uno y no información general. Incluso, con las familias de índices no se puede abarcar la totalidad de índices que existen en la literatura.

Warrens (2008) también estudió los valores mínimos y máximos y la linealidad de las familias de índices y las estructuras de las matrices de datos y las de

disimilitud, ya que si se cumplen ciertos requisitos se puede garantizar la euclideanidad. También abarca el estudio de índices multivariados lo cual escapa al objetivo de este trabajo. Concluye su trabajo afirmando que para un gran número de coeficientes de similitud se estudiaron varias propiedades y siete de ellos se destacan como los coeficientes más recomendables. Los mismos son: Jaccard, Sorensen, Braun-Blanquet, Russell-Rao, Loevinger, Sokal & Michener y Cohen.

La reseña bibliográfica de antecedentes realizada para esta Tesis muestra que la pregunta de Hubálek en 1982 - ¿cuál coeficiente de similitud (o disimilitud) es el mejor? - todavía no ha sido respondida. Este problema quizás no tenga una solución única, sino que debería ser estudiado en cada área del conocimiento y con cada tipo de datos. En Ecología de Comunidades por ejemplo, aún no existe una respuesta, en ésta área suelen presentarse gran cantidad de ceros, es decir una muy alta proporción de ausencias conjuntas, lo cual lleva a algunos índices a valores ajenos a la realidad. Por eso Legendre & Legendre (1998) recomiendan en esta área la aplicación de índices que no tienen en cuenta las ausencias conjuntas. Marti Anderson (2001a) recomienda el índice de disimilitud de Bray-Curtis (cuya contrapartida para datos binarios es el índice de Sorensen) con buenas propiedades ecológicas, sin embargo, es una semi métrica y no fueron estudiadas en forma general sus propiedades. Anderson también declara que este índice no puede ser considerado una distancia, como la distancia euclídea, porque no cumple con la propiedad de desigualdad triangular. Por lo tanto, no se conoce si la gráfica de un conjunto de datos en un espacio euclídeo respeta las distancias originales medidas con dicha disimilitud. Concluye su comentario afirmando que han sido descritas en la literatura más de sesenta medidas de similitud o disimilitud, pero con muy pocas comparaciones reales de sus comportamientos con diferentes tipos de datos ecológicos (Lamont & Grant 1979, Legendre & Legendre 1998). Esta es sin duda un área que necesita una profunda investigación.

## OBJETIVO

***El objetivo de este trabajo es evaluar un conjunto de propiedades importantes en un amplio grupo de índices de similitud o disimilitud para datos binarios estableciendo pautas de regularidad en su comportamiento e identificando cuáles serían recomendables en distintos escenarios de datos de vegetación.***

### Objetivos específicos

- 1) Realizar una profunda revisión bibliográfica que permita identificar un amplio espectro de índices de similitud y/o disimilitud, utilizados con frecuencia en Ecología.
- 2) Identificar un conjunto de propiedades consideradas de importancia para reflejar los parecidos entre unidades.
- 3) Definir las estrategias de comparación de los índices según la propiedad a estudiar.
- 4) Realizar estudios comparativos en forma analítica o por simulación según la propiedad a estudiar.
- 5) Establecer comparaciones entre los índices según cumplan o no esas propiedades definidas.
- 6) Presentar una síntesis de los resultados obtenidos, en cuadros comparativos completos de los índices, que faciliten a usuarios de distintas áreas y en especial de Ecología de comunidades, la elección de las mejores alternativas, en función de sus propiedades y en vista de los objetivos específicos de cada análisis.
- 7) Comparar los índices que resulten con mejores propiedades en conjuntos de datos de vegetación reales y simulados, con distintas particularidades, y con el propósito de decidir cuáles son los adecuados en este contexto.

El contenido de esta Tesis está estructurado en cinco capítulos incluyendo el que corresponde a esta introducción.

El Capítulo 2 de Metodología se ha dedicado a especificar la notación adoptada, las definiciones del conjunto de índices seleccionados, las propiedades a estudiar sobre cada uno de ellos y un apartado donde se explica la metodología utilizada para realizar la comprobación de las propiedades.

El Capítulo 3 muestra los resultados realizando en cada propiedad una discusión que incorpora los resultados obtenidos por otros autores.

El Capítulo 4 sintetiza las conclusiones obtenidas en el Capítulo 3 comparando la eficiencia de los setenta y siete índices estudiados. En este capítulo se deducen los mejores índices de esta colección.

El Capítulo 5 utiliza los mejores índices encontrados para realizar un análisis de ordenamiento sobre datos reales de vegetación y datos simulados, a fin de comprobar su eficacia en el análisis de datos y la consistencia entre sus resultados.

La Tesis adjunta anexos con resultados adicionales detallados y la documentación de los programas utilizados.

# CAPÍTULO 2

## METODOLOGÍA

En este Capítulo se definen una serie de setenta y siete índices que han sido propuestos en la literatura para evaluar la similitud o disimilitud entre objetos sobre los cuales se miden variables dicotómicas.

También se explicita el conjunto de quince propiedades que son evaluadas sobre cada uno de ellos, especificando características sobre su importancia para el análisis de datos.

Finalmente se detalla cómo se realiza el análisis de cada una de estas propiedades sobre los índices elegidos. Algunas de ellas son verificables simplemente con cálculos algebraicos mientras que para otras se requieren simulaciones de datos y cálculos más complejos.

### 2. 1. Notación

Sean  $\mathbf{X}_j$ ,  $j=1, \dots, r$  vectores de dimensión  $1 \times n$  que representan unidades de observación multivariadas que corresponden a  $r$  realizaciones de un vector aleatorio, cuyas componentes identifican a datos binarios codificados con 0 y 1, donde el 0 representa la ausencia y el 1 la presencia de diferentes atributos, es decir:

$$\mathbf{X}_j = (0 \ 1 \ 0 \ \dots \dots \dots \ 1 \ 0 \ 0) , j=1, \dots, r$$

El conjunto total de observaciones  $\mathbf{X}_j$  se indica con  $\mathbf{X}_{r \times n}$ , matriz cuyas filas corresponden a cada unidad de observación.

Una forma conveniente de resumir la información sobre el parecido entre dos de tales unidades  $\mathbf{X}_j$  y  $\mathbf{X}_k$  es a través de cuatro cantidades  $a$ ,  $b$ ,  $c$  y  $d$  donde:

$a$ : es el número de atributos que  $\mathbf{X}_j$  y  $\mathbf{X}_k$  tienen en común (presencia conjunta)

$b$ : es el número de atributos presentes en  $\mathbf{X}_j$  solamente

$c$ : es el número de atributos presentes en  $\mathbf{X}_k$  solamente

$d$ : es el número de atributos ausentes en  $\mathbf{X}_j$  y  $\mathbf{X}_k$  (ausencia conjunta)

Estas cuatro cantidades suelen presentarse en forma de tabla de contingencias de 2x2 (Tabla 1) y son utilizados para definir indicadores que cuantifiquen su parecido (o diferencia).

Tabla 1: Comparación de dos unidades experimentales  $\mathbf{X}_j$  y  $\mathbf{X}_k$  sobre la base de la presencia/ausencia de  $n$  características

		Obs. $\mathbf{X}_k$		
		presente	ausente	
Obs. $\mathbf{X}_j$	presente	a	b	a+b
	ausente	c	d	c+d
		a+c	b+d	n=a+b+c+d

En la Tabla 1, el valor esperado de  $a$  bajo el supuesto de independencia entre los vectores aleatorios  $\mathbf{X}_j$  y  $\mathbf{X}_k$  es:

$$E(a) = a' = \frac{(a+b)(a+c)}{n}$$

Los coeficientes de similitud (o de disimilitud) se construyen combinando las cantidades  $a$ ,  $b$ ,  $c$ ,  $d$  y  $a'$ . Se indica con  $A_i(j, k)^{(a,b,c,d)}$  la similitud de un índice genérico  $i$  entre los vectores  $\mathbf{X}_j$  y  $\mathbf{X}_k$ , la cual depende de las cantidades  $a$ ,  $b$ ,  $c$ ,  $d$ . En numerosas aplicaciones de análisis multivariados estos indicadores se presentan en matrices cuadradas y simétricas que contienen los parecidos entre todos los pares de unidades de análisis disponibles.

La matriz de los indicadores de similitud se indica con  $\mathbf{A}_{rxr} = \{A_i(j, k)^{(a,b,c,d)}\}$  y la de disimilitud con  $\mathbf{D}_{rxr} = \{D_i(j, k)^{(a,b,c,d)}\}$ , para un índice genérico  $i$  y para  $j, k=1, \dots, r$ .

La principal distinción que se hace entre los coeficientes es si consideran en su fórmula las ausencias conjuntas, es decir, la cantidad  $d$ . Si una variable binaria se define como la presencia o ausencia de un atributo, entonces la cantidad  $d$  refleja las coincidencias negativas, lo cual en general no contribuye a la similitud. Esto ocurre en Ecología de comunidades, donde se registran las presencias o ausencias de especies vegetales en un conjunto de sitios y la ausencia conjunta

de algunas de ellas en dos sitios no indica que los mismos sean parecidos. Por el contrario si la información que se registra es la presencia de flores amarillas o rojas en una planta, ambas, las presencias y ausencias conjuntas aportan información valiosa para el estudio.

## 2. 2. Definiciones

A continuación se encuentra la lista de todos los coeficientes de similitud considerados en esta Tesis, ordenados alfabéticamente por autores y la notación con la cual se mencionan en este trabajo. Cada índice se identifica por el nombre del autor y el año en que fue propuesto. En caso de que un autor (o varios) haya propuesto más de un índice éstos se indican con un subíndice numérico. La mayoría de los índices aparecen en el trabajo de Hubálek (1982) y se respetan los nombres que allí figuran. Se completa la lista con índices que publica Warrens (2008) en su trabajo, presentando cinco coeficientes de los cuales no se ha encontrado ninguna información en la bibliografía consultada. En este trabajo se respeta su notación y aparecen como Anónimos. Se agregan luego algunos índices adicionales encontrados en el paquete Simba de R. Información adicional sobre los mismos se incluye de Rodríguez-Salazar et al. (2001).

En este trabajo a los índices que fueron desarrollados simultáneamente por dos o más autores se los identifica con el nombre del primer autor que figura en la lista alfabética. Los cinco índices desarrollados por Sokal & Sneath son numerados de acuerdo al orden de aparición en el trabajo de Hubálek (1982).

En la enumeración de los índices que aparece a continuación, por simplicidad, se utiliza una nomenclatura simplificada:

$$A_i = A_i(j, k) \text{ }^{(a,b,c,d)}$$

### Anónimo\_1

$$A_1 = \frac{2a - (b + c)}{2a + (b + c)}$$

### Anónimo\_2

$$A_2 = \frac{2d}{2d + b + c}$$

**Anónimo\_3**

$$A_3 = \frac{2d - (b + c)}{2d + (b + c)}$$

**Anónimo\_4**

$$A_4 = \frac{4ad}{4ad + (a + d)(b + c)}$$

**Anónimo\_5**

$$A_5 = \frac{ad - bc}{\min\{(a + b)(a + c), (c + d)(b + d)\}}$$

**Baroni-Urbani & Buser\_1 (1976)**

$$A_6 = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

**Baroni-Urbani & Buser\_2 (1976)**

$$A_7 = \frac{\sqrt{ad} + a - b - c}{\sqrt{ad} + a + b + c}$$

**Braun-Blanquet (1932)**

$$A_8 = \frac{a}{\max\{(a + b), (a + c)\}}$$

es el mínimo entre Dice\_1 y Dice\_2

**Clement (1976)**

$$A_9 = \frac{a(c + d)}{a + b} + \frac{d(a + b)}{c + d}$$

**Clifford & Stephenson (1975)**

$$A_{10} = \frac{\chi^2}{n} = \frac{A_{54}}{n} = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

**Cohen (1960)**

$$A_{11} = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}$$

coincide con la media armónica de Cole\_3 y Cole\_4

**Cole\_1 (1949)**

$$A_{12} = \frac{\sqrt{2}(ad - bc)}{\sqrt{(ad - bc)^2 + (a + b)(c + d)(a + c)(b + d)}}$$

**Cole\_2 (1949, 1957)**

$$A_{13} = \frac{\phi}{\phi_{\text{máx.}}} = \frac{A_{75}}{(A_{75})_{\text{máx.}}} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} \phi_{\text{máx.}}$$

donde  $\Phi$  es el índice de Yule\_1, y  $\Phi_{\text{máx.}}$  es el mayor valor posible de  $\Phi$  consistente con los totales marginales

**Cole\_3 (1949)**

$$A_{14} = \frac{ad - bc}{(a + b)(b + d)}$$

también denominado coeficiente de asociación interespecífica

**Cole\_4 (1949)**

$$A_{15} = \frac{ad - bc}{(a + c)(c + d)}$$

**Dennis, en Holliday et al. (2002), Ellis et al. (1993)**

$$A_{16} = \frac{ad - bc}{\sqrt{(a + b + c + d)(a + b)(a + c)}}$$

**Dice\_1 (1945), Simpson\_2 (1960), Shi (1993)**

$$A_{17} = \frac{a}{a + b}$$

**Dice\_2 (1945), Ruggiero et al. (1998), Koleff et al. (2003)**

$$A_{18} = \frac{a}{a + c}$$

**Dice\_3 (1945)**

$$A_{19} = \frac{2a}{(a + b)(a + c)}$$

también conocido como índice de incidencia

**Digby (1983)**

$$A_{20} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}}$$

**Doolittle (1885), Pearson (1926)**

$$A_{21} = A_{75}^2 = \phi^2 = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

**Eyraud (1936)**

$$A_{22} = \frac{a - (a + b)(a + c)}{(a + b)(c + d)(a + c)(b + d)}$$

**Fager & McGowan (1963)**

$$A_{23} = \frac{a}{\sqrt{(a + b)(a + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(a + b), (a + c)\}}}$$

**Faith (1983)**

$$A_{24} = \frac{a + \frac{d}{2}}{n} = \frac{a + \frac{d}{2}}{a + b + c + d}$$

**Fleiss (1975)**

$$A_{25} = \frac{(ad - bc)[(a + b)(b + d) + (a + c)(c + d)]}{2(a + b)(a + c)(b + d)(c + d)}$$

**Forbes\_1 (1907)**

$$A_{26} = \frac{a}{a'} = \frac{a}{\frac{(a+b)(a+c)}{n}} = \frac{n * a}{(a+b)(a+c)} = \frac{(a+b+c+d)a}{(a+b)(a+c)}$$

**Forbes\_2 (1925)**

$$A_{27} = \frac{a - a'}{a_{\text{máx.}} - a'} = \frac{n * a - (a+b)(a+c)}{n * \text{máx.}\{(a+b), (a+c)\} - (a+b)(a+c)}$$

**Forbes\_3 (en Cole 1949)**

$$A_{28} = \frac{ad - bc}{(a+b)(a+c)}$$

**Fossum, en Holliday et al. (2002), Ellis et al. (1993)**

$$A_{29} = \frac{n(a - \frac{1}{2})^2}{(a+b)(a+c)} = \frac{(a+b+c+d)(a - \frac{1}{2})^2}{(a+b)(a+c)}$$

**Gilbert & Wells (1966)**

$$A_{30} = \log(a) - \log(n) - \log\left(\frac{a+b}{n}\right) - \log\left(\frac{a+c}{n}\right)$$

no definido para a=0

**Goodall (1967); Austin & Colwell (1977)**

$$A_{31} = \frac{2}{\pi} \arcsin \sqrt{A_{64}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{n}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+c+d}}$$

**Goodman & Kruskal (1954)**

$$A_{32} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

**Hamann (1961), Hubert (1977)**

$$A_{33} = \frac{(a+d) - (b+c)}{n} = \frac{a+d-b-c}{a+d+b+c}$$

en psicología es conocido como índice G (Holley & Guilford 1964)

**Harris & Lahey (1978)**

$$A_{34} = \frac{a(2d + b + c)}{2(a + b + c)} + \frac{d(2a + b + c)}{2(d + b + c)}$$

**Hawkins & Dotson (1968)**

$$A_{35} = \frac{1}{2} \left( \frac{a}{a + b + c} + \frac{d}{d + b + c} \right)$$

**Hurlbert\_1 (1969)**

$$A_{36} = \pm \sqrt{\frac{\chi^2}{\chi_{m\acute{a}x}^2}} = \pm \sqrt{\frac{A_{54}}{(A_{54})_{m\acute{a}x}}} = \pm \sqrt{\frac{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}}{(A_{54})_{m\acute{a}x}}}$$

donde el signo  $\pm$  se corresponde con el signo de  $(ad-bc)$  y  $\chi_{m\acute{a}x}^2$  es el mayor valor posible de  $\chi^2$  consistente con los totales marginales

**Hurlbert\_2 (1969)**

$$A_{37} = \pm \sqrt{\frac{\chi^2 - \chi_{m\acute{i}n}^2}{\chi_{m\acute{a}x}^2 - \chi_{m\acute{i}n}^2}} = \pm \sqrt{\frac{A_{54} - (A_{54})_{m\acute{i}n}}{(A_{54})_{m\acute{a}x} - (A_{54})_{m\acute{i}n}}}$$

$$= \pm \sqrt{\frac{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} - (A_{54})_{m\acute{i}n}}{(A_{54})_{m\acute{a}x} - (A_{54})_{m\acute{i}n}}}$$

donde el signo  $\pm$  se corresponde con el signo de  $(ad-bc)$  y  $\chi_{m\acute{a}x}^2$  y  $\chi_{m\acute{i}n}^2$  son el mayor y el menor valor posible de  $\chi^2$  consistente con los totales marginales respectivamente

**Jaccard (1901), Sneath (1957), Iversen (1954), Agrell (1945)**

$$A_{38} = \frac{a}{a + b + c}$$

tambi3n conocido como coeficiente de comunidad

**Johnson\_1 (1967)**

$$A_{39} = \frac{a}{a + b} + \frac{a}{a + c}$$

**Johnson\_2 (1971)**

$$A_{40} = \frac{a}{2b}$$

**Kendall (en Podani 1997)**

$$A_{41}$$

$$= \frac{2(ad - bc)}{\sqrt{(n^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)(n^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

desarrollado por Podani para datos binarios

**Kuder & Richardson (1937), Cronbach (1951)**

$$A_{42} = \frac{4(ad - bc)}{(a + b)(c + d) + (a + c)(b + d) + 2(ad - bc)}$$

**Kulczynski\_1 (1927), Driver & Kroeber (1932)**

$$A_{43} = \frac{a}{b + c}$$

**Kulczynski\_2 (1927), Driver & Kroeber (1932)**

$$A_{44} = \frac{1}{2} \left( \frac{a}{a + b} + \frac{a}{a + c} \right)$$

coincide con la media aritmética de Dice\_1 y Dice\_2

**Lamont & Grant (1979)**

$$A_{45} = \frac{a}{2a + b + c}$$

**Legendre & Legendre (1998)**

$$A_{46} = \frac{3a}{3a + b + c}$$

**Loevinger (1947, 1948), Benini (1901)**

$$A_{47} = \frac{ad - bc}{\min. \{(a + b)(b + d), (a + c)(c + d)\}}$$

coincide con el máximo de Cole\_3 y Cole\_4

**Maxwell & Pilliner (1968)**

$$A_{48} = \frac{2(ad - bc)}{(a + b)(c + d) + (a + c)(b + d)}$$

**McConnaughey (1964)**

$$A_{49} = \frac{a^2 - bc}{(a + b)(a + c)}$$

**Michael (1920)**

$$A_{50} = \frac{4(ad - bc)}{(a + d)^2 + (b + c)^2}$$

**Mountford (1962)**

$$A_{51} = \frac{a}{\frac{1}{2}(ab + ac) + bc}$$

**Ochiai (1957), Driver & Kroeber (1932), Barkman (1958)**

$$A_{52} = \frac{a}{\sqrt{(a + b)(a + c)}}$$

coincide con la media geométrica de Dice\_1 y Dice\_2

**Pearson & Heron (1913)**

$$A_{53} = \cos \frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

también conocido como coeficiente de correlación tetraclórico

**Pearson\_1 (1905), en Yule & Kendall (1950)**

$$A_{54} = \chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

**Pearson\_2 (1905), en Yule & Kendall (1950)**

$$A_{55} = + \sqrt{\frac{\chi^2}{n + \chi^2}} = + \sqrt{\frac{A_{54}}{n + A_{54}}} = + \sqrt{\frac{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}}{n + \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}}$$

**Peirce\_1 (1884)**

$$A_{56} = \frac{ab + bc}{ab + 2bc + cd}$$

**Peirce\_2 (1884)**

$$A_{57} = \frac{ad - bc}{(a + c)(b + d)}$$

**Peirce\_3 (1884)**

$$A_{58} = \frac{ad - bc}{(a + b)(c + d)}$$

**Rogers & Tanimoto (1960)**

$$A_{59} = \frac{a + d}{a + 2(b + c) + d}$$

**Rogot & Goldberg (1966)**

$$A_{60} = \frac{a}{2a + b + c} + \frac{d}{2d + b + c}$$

coincide con la media aritmética de Anónimo\_1 y Sorensen

**Russell & Rao (1940)**

$$A_{61} = \frac{a}{n} = \frac{a}{a + b + c + d}$$

**Scott (1955)**

$$A_{62} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(b + c + 2d)}$$

**Simpson (1943)**

$$A_{63} = \frac{a}{\min.\{(a + b), (a + c)\}}$$

es el máximo entre Dice\_1 y Dice\_2

**Sokal & Michener (1958)**

$$A_{64} = \frac{a + d}{n} = \frac{a + d}{a + b + c + d}$$

coeficiente de igualdad simple (SMC, por sus siglas en ingles *simple matching coefficient*)

**Sokal & Sneath\_1 (1963)**

$$A_{65} = \frac{a}{a + 2(b + c)}$$

**Sokal & Sneath\_2 (1963)**

$$A_{66} = \frac{1}{4} \left( \frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{c + d} + \frac{d}{b + d} \right)$$

**Sokal & Sneath\_3 (1963)**

$$A_{67} = \frac{a + d}{b + c}$$

**Sokal & Sneath\_4 (1963)**

$$A_{68} = \frac{a + d}{a + \frac{1}{2}(b + c) + d}$$

**Sokal & Sneath\_5 (1963)**

$$A_{69} = \frac{ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

coincide con el cuadrado de la media geométrica de  $\frac{a}{a+b}$ ,  $\frac{a}{a+c}$ ,  $\frac{d}{c+d}$  y  $\frac{d}{b+d}$

**Sorensen (1948), Dice (1945), Gleason (1920), Czekanowski (1913)**

$$A_{70} = \frac{2a}{2a + b + c} = \frac{a}{a + \frac{1}{2}(b + c)}$$

coincide con la media armónica de Dice\_1 y Dice\_2

**Sorgenfrei (1959), Van der Maarel (1966)**

$$A_{71} = \frac{a^2}{(a+b)(a+c)}$$

es el producto de Dice\_1 y Dice\_2

**Stiles (1961)**

$$A_{72} = \log \frac{n \left( |ad - bc| - \frac{n}{2} \right)^2}{(a+b)(a+c)(b+d)(c+d)}$$

**Tarwid (1960)**

$$A_{73} = \frac{a - a'}{a + a'} = \frac{n * a - (a+b)(a+c)}{n * a + (a+b)(a+c)}$$

**Tschuproff (en Rodriguez-Salazar et al.2001)**

$$A_{74} = + \sqrt{\frac{\chi^2}{n}} = + \sqrt{\frac{A_{54}}{n}} = + \sqrt{\frac{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{n}}$$

$$= + \sqrt{\frac{(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

**Yule\_1 (1912), Pearson & Heron (1913)**

$$A_{75} = \phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

usualmente denominado con  $\Phi$ , coincide con el coeficiente de correlación de Pearson aplicado a datos binarios, coincide con la media geométrica de Cole\_3 y Cole\_4

**Yule\_2 (1900), Yule & Kendall (1950)**

$$A_{76} = \frac{ad - bc}{ad + bc}$$

coincide con el coeficiente de Goodman & Kruskal para datos binarios, como fue desarrollado por Podani (1997)

### Yule\_3 (1912)

$$A_{77} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

A partir de las medidas de similitud se pueden derivar medidas de disimilitud. Si la similitud  $A_i$  varía en el rango  $[0,1]$ , se pueden definir varios coeficientes de disimilitud  $D_i$ :

$$D1_i = 1 - A_i \quad D2_i = \sqrt{1 - A_i} \quad D3_i = \sqrt{2(1 - A_i)}$$

Los coeficientes de similitud  $A_i$  que varían en el rango  $[-1,1]$  pueden ser convertidos a coeficientes de disimilitud (Clifford & Stephenson 1975) de la siguiente forma:

$$D1_i = \frac{1}{2}(1 - A_i) \quad D2_i = \sqrt{\frac{1}{2}(1 - A_i)} \quad D3_i = \sqrt{1 - A_i}$$

Algunas de las propiedades consideradas en este trabajo se evalúan también para estas tres variantes de medidas de disimilitud derivadas de las similitudes.

Existen otras medidas de disimilitud que podrían definirse a partir de las similitudes (Legendre & Legendre 1998, Demey 2011), pero que no serán tenidas en cuenta en este trabajo, como por ejemplo:

$$\sqrt{1 - A_i^2} \quad -\log(A_i) \quad \frac{1}{A_i} - 1 \quad \frac{1}{A_i - 1}$$

En su definición los coeficientes de similitud tienen generalmente en el numerador los términos:  $a$ ,  $a+d$  ó  $ad-bc$ . Mientras que el término  $a$  en el numerador es obvio y evidente por sí mismo, existe alguna controversia con respecto a la inclusión de las ausencias conjuntas en el término  $a+d$  (Sokal & Sneath 1963, Jardine & Sibson 1971, Goodall 1973, Sneath & Sokal 1973, Clifford & Stephenson 1975). Los coeficientes con  $a+d$  en el numerador han sido exitosamente aplicados en Taxonomía. Por otro lado, su uso en Ecología está limitado por la presencia de valores generalmente altos de  $d$ , lo cual por supuesto, puede afectar grandemente la medida de similitud bajo ciertas circunstancias. Legendre & Legendre (1998) llaman a los índices que tienen en

cuenta solamente el término  $a$  coeficientes asimétricos y simétricos cuando también tienen en cuenta los dobles ceros. En este trabajo no se utiliza esa definición ya que una de las propiedades a estudiar es la simetría de los índices y puede dar lugar a confusión.

### **2. 3. Propiedades de los índices de similitud y/o disimilitud**

En este apartado se detallan en total quince propiedades que se estudian en este trabajo. Trece de ellas son consideradas de interés por Hubálek (1982) aunque en su trabajo solo fueron estudiadas sobre un subconjunto de los setenta y siete índices considerados en esta tesis, incluso hay propiedades que este autor define pero que no fueron estudiadas en ningún caso particular. A estas trece se agregan dos propiedades más, la indeterminancia y la euclideanidad, haciendo un total de quince propiedades.

#### **Propiedad 1: Existencia de límites convencionales**

Un índice de similitud  $A_i$  tiene límites convencionales si sus valores están dentro de los siguientes rangos:

$$-1 \leq A_i \leq 1$$

$$0 \leq A_i \leq 1$$

En ambos casos, el valor máximo de  $A_i$  dentro de cada rango significa el mayor grado de similitud entre las unidades comparadas.

Los índices que satisfacen la primer condición, es decir que varían entre -1 y 1 son llamados centrados por Dagnelie (1960), mientras que en la otra situación se los llama no centrados.

Para los índices que varían en el intervalo  $[0,1]$ , el valor de 0 corresponde a “similitud nula” mientras que el valor de 1 indica “similitud completa”. Para los índices que varían en el intervalo  $[-1,1]$ , un valor de 1, entre dos unidades que se comparan, indica que las mismas son similares en todos los aspectos, mientras que un valor de -1 señala que son disímiles en todos los aspectos. Los valores nulos del coeficiente indican que el par de unidades comparadas pueden ser consideradas como igualmente similares o bien como igualmente disímiles (Clifford y Stephenson 1975).

## Propiedad 2: Maximal y Minimal

Un coeficiente de similitud se dice maximal si:

$$\underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, j)^{(a,b,c,d)}$$

De este modo, si:

$A_i(j, k) > A_i(l, m)$  luego se dice que  $X_j$  y  $X_k$  son más similares que  $X_l$  y  $X_m$

De la misma manera, un coeficiente de disimilitud se dice minimal si:

$$\underbrace{\text{mín.}}_{j,k=1,\dots,r} D_i(j, k)^{(a,b,c,d)} = D_i(j, j)^{(a,b,c,d)}$$

De este modo, si:

$D_i(j, k) > D_i(l, m)$  luego se dice que  $X_j$  y  $X_k$  son más disímiles que  $X_l$  y  $X_m$

## Propiedad 3: Existencia de valores mínimos y máximos

Un coeficiente de similitud debería alcanzar su valor mínimo cuando  $a=d=0$  y su valor máximo cuando  $b=c=0$  (Baroni-Urbani & Buser 1976), es decir:

$$\text{a) } \underbrace{\text{mín.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a=0,b,c,d=0)}$$

$$\text{b) } \underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)}$$

## Propiedad 4: Asociación completa o absoluta

Esta propiedad solo se define para aquellos índices de similitud que alcanzan su valor máximo cuando  $b=c=0$  (Propiedad 3b).

$A_i$  se dice un coeficiente de asociación completa (Kendall & Stuart 1967, Pielou 1969) si también alcanza su valor máximo cuando  $b=0$  o  $c=0$ :

$$\underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} = A_i(j, k)^{(a,b=0,c,d)} \quad \text{y}$$

$$\underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} = A_i(j, k)^{(a,b,c=0,d)} \quad ,$$

y  $A_i$  se dice un coeficiente de asociación absoluta si:

$$\underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} > A_i(j, k)^{(a,b=0,c,d)} \quad y$$

$$\underbrace{\text{máx.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} > A_i(j, k)^{(a,b,c=0,d)}$$

### Propiedad 5: Asociación nula cuando el valor de a=0

Un índice cumple con la condición de asociación nula si alcanza el valor mínimo cuando el valor de  $a$  es cero (Baroni-Urbani & Buser 1976), es decir:

$$\underbrace{\text{mín.}}_{j,k=1,\dots,r} A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a=0,b,c,d)}$$

Esta es una condición más estricta que la propiedad 3a).

### Propiedad 6: Simetría

Se dice que un índice  $A_i$  es simétrico si cumple:

$$A_i(j, k)^{(a,b,c,d)} = A_i(k, j)^{(a,b,c,d)} \quad \text{para todo } j, k=1,\dots, r$$

De la misma manera, un índice de disimilitud es simétrico si:

$$D_i(j, k)^{(a,b,c,d)} = D_i(k, j)^{(a,b,c,d)} \quad \text{para todo } j, k=1,\dots, r$$

Equivalentemente, si un índice  $A_i$  o  $D_i$  es invariante frente al intercambio entre “b” y “c”, cumple la propiedad de simetría.

### Propiedad 7: Monotonicidad respecto de la estadística $\chi^2$

Se dice que  $A_i$  cumple la condición de monotonicidad si es una función monótona no decreciente de la estadística  $\chi^2$ , la estadística que evalúa independencia en una tabla de contingencias de  $2 \times 2$ . En esta definición se considera que la estadística de prueba del test  $\chi^2$  de independencia es la mejor y más sencilla medida de fuerza de asociación entre dos objetos o unidades muestrales con datos cualitativos con dos estados solamente. Se asume que con una diferencia creciente de “a” con el valor esperado  $a'$  en la dirección positiva, el valor de un índice de asociación también debería incrementarse para valores fijos de  $n$ . Debido a que  $\chi^2$  no discrimina entre valores positivos o negativos de la desviación  $(a - a')$ , cuando se evalúa la monotonicidad de aquellos índices que

toman valores en el intervalo  $[-1,1]$ , el rango de los valores debería ser particionado en dos subgrupos, uno para  $(ad - bc) \leq 0$  y otro para  $(ad - bc) > 0$ . Para verificar la monotonicidad se computa la correlación de Spearman o Kendall entre  $A_i$  y  $\chi^2$ .

Hay que observar que Hubálek recomienda solamente utilizar el coeficiente momento-producto de Pearson entre  $A_i$  y  $\sqrt{\chi^2}$ , pero este no resulta adecuado ya que este coeficiente mide la relación lineal entre ambas, que es más fuerte que la monotonicidad. Por otro lado, no es necesario considerar la raíz cuadrada, ya que esta es monótona con respecto a  $\chi^2$ .

### Propiedad 8: Linealidad

Un índice  $A_i$  cumple con la condición de linealidad si:

$$A_i(j, k)^{(a+1, b-1, c-1, d+1)} - A_i(j, k)^{(a, b, c, d)} = A_i(j, k)^{(a+2, b-2, c-2, d+2)} - A_i(j, k)^{(a+1, b-1, c-1, d+1)}$$

$\forall a, b, c, d$  tal que  $a+b+c+d=n$  y cuyas tablas de asociación tengan los mismos totales marginales

Utilizando una notación reducida la propiedad se enuncia:

$$A_i(j, k)^{(a+1)} - A_i(j, k)^{(a)} = A_i(j, k)^{(a+2)} - A_i(j, k)^{(a+1)}$$

Si no se cumple la condición de linealidad, el índice es cóncavo o convexo.

Un índice  $A_i$  se dice convexo si:

$$A_i(j, k)^{(a+1)} - A_i(j, k)^{(a)} < A_i(j, k)^{(a+2)} - A_i(j, k)^{(a+1)}$$

Un índice  $A_i$  se dice cóncavo si:

$$A_i(j, k)^{(a+1)} - A_i(j, k)^{(a)} > A_i(j, k)^{(a+2)} - A_i(j, k)^{(a+1)}$$

La comprobación de esta propiedad requiere la realización de múltiples comparaciones que consideran escenarios particulares donde los totales marginales de las tablas de contingencias de  $2 \times 2$  son fijos, para cualquier  $n$ . Distintos autores establecen situaciones particulares diferentes. En particular Hubálek plantea cuatro variantes computadas para cada coeficiente  $A_i^{(a,b,c,d)}$ :  $A_i(0,40,40,20)$ ,  $A_i(1,39,39,21)$ ,  $A_i(2,38,38,22)$  y  $A_i(3,37,37,23)$ .

Otros autores como Gower & Legendre (1986) y Hajdu (1981) proponen estrategias más generales de esta propiedad de linealidad, planteando escenarios con más variantes que las presentadas por Hubálek (1982).

Hajdu (1981) desarrolló las series de comparaciones de casos ordenados (sus siglas en inglés OCCAS *ordered comparison case series*). Cada serie se construye con datos cuantitativos artificiales para comparar dos unidades experimentales con cambios lineales desde un caso al siguiente, proponiendo verificar si el efecto de cambio que se produce en los coeficientes es también lineal.

Gower y Legendre (1986) extienden el esquema de Hajdu para cubrir variables binarias. A tal efecto consideran un rango de  $t$  valores igualmente espaciados para  $a$ ,  $b$ ,  $c$  y  $d$  y definen las diferencias (utilizando la notación reducida):

$$\delta_h = |A_i(j, k)^{(h+1)} - A_i(j, k)^{(h)}|, \text{ donde } h=1,2,\dots,(t-1)$$

Luego, las desviaciones de la linealidad, llamadas no-linealidad, se resumen o condensan en el desvío estándar de  $\delta_h$  y en la media aritmética de  $\delta_h$  denominada resolución. El cociente entre ambas medidas define un coeficiente de variación (multiplicado por 100), el cual debería ser cercano a cero para el cumplimiento de la propiedad.

$$CV = \frac{\text{no - linealidad}}{\text{resolucion}} * 100 = \frac{\text{desvío estándar}(\delta_h)}{\text{media aritmética}(\delta_h)} * 100$$

Como señalan Gower & Legendre (1986) una alta resolución es una propiedad deseable en Análisis Cluster y principalmente en los métodos del promedio y del centroide. Hajdu (1981) afirma que un coeficiente con una alta resolución forma grupos más estables y son menos probables de ser modificados ante pequeños cambios en los datos. Por su parte, los índices que cumplen la condición de linealidad son apropiados para Escalamiento métrico, como por ejemplo Análisis de Coordenadas Principales, como fue expresado en la Introducción. Si se utiliza Escalamiento no métrico, como *Non Metric Multidimensional Scaling*, la linealidad podría mejorar la bondad del del procedimiento reduciendo medidas de “stress”, pero el ordenamiento resultante no se vería afectado.

### Propiedad 9: Insensibilidad al valor de $a=0$

Un índice  $A_i$  se dice insensible al valor de  $a=0$  si los cambios producidos desde un valor de  $A_i(j,k)^{(a,b,c,d)}$  con  $a=0$ , frente a cambios unitarios de  $a$ , son graduales. Es decir, si no hay cambios bruscos consecutivamente entre:  $A_i(j,k)^{(a=0,b,c,d)}$ ,  $A_i(j,k)^{(a=1,b-1,c-1,d+1)}$ ,  $A_i(j,k)^{(a=2,b-2,c-2,d+2)}$ ,  $A_i(j,k)^{(a=3,b-3,c-3,d+3)}$ , etc. Nótese que al cambiar “a” cambian “b”, “c” y “d” de modo de mantener los totales marginales.

### Propiedad 10: Asociación positiva o negativa

Un índice de similitud es capaz de reconocer entre asociación positiva o negativa cuando su valor cambia según la relación entre frecuencias observadas y esperadas en la Tabla 1 (Yule 1912, Yule & Kendall 1950).

Recordando que bajo independencia:

$$E(a) = a' = \frac{(a+b)(a+c)}{n}$$

un índice  $A_i$  cumple con la propiedad de asociación positiva si para cada  $a'$  fijo y manteniendo los totales marginales es:

$$A_i(j,k)^{(a_1,b,c,d)} < A_i(j,k)^{(a_2,b,c,d)} \quad \text{para } a_1 < a' < a_2 \quad \forall j, k=1, \dots, r$$

En caso contrario:

$$A_i(j,k)^{(a_1,b,c,d)} > A_i(j,k)^{(a_2,b,c,d)} \quad \text{para } a_1 < a' < a_2 \quad \forall j, k=1, \dots, r$$

se dice que la asociación es negativa.

### Propiedad 11: Distribuciones en el muestreo

Los índices de similitud y/o disimilitud son variables aleatorias que dependen de las frecuencias absolutas de la Tabla de contingencia 1. Sus distribuciones en el muestreo son desconocidas. En esta tesis se estudia si es posible ajustar alguna distribución conocida a las distribuciones en el muestreo de los setenta y siete índices de similitud y de las tres disimilitudes derivadas, bajo simulación de los vectores de datos binarios. En particular se enfoca el ajuste a la distribución normal al menos en cuanto a asimetría y curtosis.

### **Propiedad 12: Insesgamiento**

Los índices deberían tener buenas propiedades de muestreo, entre ellas el insesgamiento. Es decir, si se toman muestras repetidas desde una población con probabilidades conocidas para  $a$ ,  $b$ ,  $c$  y  $d$  y se calculan con cada una los valores de  $A_i$ , su promedio debería aproximarse al valor poblacional. Se define el sesgo:

$$\text{sesgo } A_i(j, k) = E(\widehat{A}_i(j, k)) - A_i(j, k)$$

### **Propiedad 13: Métrica**

Se define disimilitud una aplicación que a cada par de elementos  $X_j$  y  $X_k$  asocia un valor real que cumple con las siguientes tres condiciones:

- 1)  $D_i(j, k) \geq 0$  para todo  $j$  y  $k$
- 2)  $D_i(j, k) = 0$  solo si  $j = k$
- 3)  $D_i(j, k) = D_i(k, j)$  para todo  $j$  y  $k$  (simetría)

Una disimilitud es una distancia o métrica si también cumple con la cuarta condición:

- 4)  $D_i(j, k) \leq D_i(j, m) + D_i(m, k)$  para todo  $j, k$  y  $m$  (desigualdad triangular)

### **Propiedad 14: Indeterminancia**

Un índice de similitud  $A_i$ , definido como fracción, se dice indeterminado si su denominador toma el valor cero para alguna variante posible de  $a$ ,  $b$ ,  $c$  y/o  $d$  (Warrens 2008).

Muchos coeficientes de similitud son definidos como fracciones, considerando en el numerador las presencias y/o ausencias conjuntas y en el denominador alguna medida de la cantidad de atributos encontrados. Como fue notado por Batagelj & Bren (1995) se le ha dado sorprendentemente poca atención a este caso de indeterminancia para algunos valores de los coeficientes para datos binarios. En este trabajo, para aquellos índices de similitud que se hacen indeterminados se utilizaron las nuevas definiciones siguiendo las recomendaciones de Warrens (2008) y Batagelj & Bren (1995), quienes re-definen apropiadamente valores en los casos críticos. Estas nuevas definiciones

se utilizaron para probar las propiedades que hicieron necesario simular matrices de datos binarios: Existencia de límites convencionales (prop.1), Maximal y Minimal (prop.2), Simetría (prop.6), Monotonicidad respecto de la estadística  $\chi^2$  (prop.7), Distribuciones en el muestreo (prop.11), Métrica (prop.13) y Euclideanidad (prop.15).

A continuación se detallan las re-definiciones de los índices de similitud dadas por Warrens (2008), que contemplan los casos en que el denominador del mismo se hace cero y que se adoptan en este trabajo.

Baroni-Urbani & Buser\_1

$$A_6 = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

$$= 0 \text{ si } d = n$$

Braun-Blanquet

$$A_8 = \frac{a}{\text{máx.}\{(a + b), (a + c)\}}$$

$$= 0 \text{ si } d = 0$$

Cohen

$$A_{11} = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

Fleiss

$$A_{25} = \frac{(ad - bc)[(a + b)(b + d) + (a + c)(c + d)]}{2(a + b)(a + c)(b + d)(c + d)}$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0 \text{ ó } (b + d) = 0 \text{ ó } (c + d) = 0$$

Goodman & Kruskal

$$A_{32} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

Hawkins & Dotson

$$A_{35} = \frac{1}{2} \left( \frac{a}{a+b+c} + \frac{d}{d+b+c} \right)$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$

Jaccard

$$A_{38} = \frac{a}{a+b+c}$$
$$= 0 \text{ si } d = n$$

Kulczynski\_2

$$A_{44} = \frac{1}{2} \left( \frac{a}{a+b} + \frac{a}{a+c} \right)$$
$$= 0 \text{ si } (a+b) = 0 \text{ ó } (a+c) = 0$$

Loevinger

$$A_{47} = \frac{ad - bc}{\text{mín.} \{ (a+b)(b+d), (a+c)(c+d) \}}$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$
$$= 0 \text{ si } (a+b) = 0 \text{ ó } (a+c) = 0 \text{ ó } (b+d) = 0 \text{ ó } (c+d) = 0$$

Maxwell & Pilliner

$$A_{48} = \frac{2(ad - bc)}{(a+b)(c+d) + (a+c)(b+d)}$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$
$$= 0 \text{ si } (a+b) = 0 \text{ ó } (a+c) = 0 \text{ ó } (b+d) = 0 \text{ ó } (c+d) = 0$$

Mc Connaughey

$$A_{49} = \frac{a^2 - bc}{(a+b)(a+c)}$$
$$= 0 \text{ si } (a+b) = 0 \text{ ó } (a+c) = 0$$

Ochiai

$$A_{52} = \frac{a}{\sqrt{(a+b)(a+c)}}$$
$$= 0 \text{ si } (a+b) = 0 \text{ ó } (a+c) = 0$$

Scott

$$A_{62} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(2d + b + c)}$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

Simpson

$$A_{63} = \frac{a}{\text{mín.}\{(a + b), (a + c)\}}$$

$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0$$

Sokal & Sneath\_2

$$A_{66} = \frac{1}{4} \left( \frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{d + c} + \frac{d}{d + b} \right)$$

$$= \frac{2a + b}{3(a + b)} \text{ si } (c + d) = 0$$

$$= \frac{2a + c}{3(a + c)} \text{ si } (b + d) = 0$$

$$= \frac{2d + b}{3(d + b)} \text{ si } (a + c) = 0$$

$$= \frac{2d + c}{3(c + b)} \text{ si } (a + b) = 0$$

$$= 0 \text{ si } b = n \text{ ó } c = n$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

Note que  $\frac{2a+b}{3(a+b)}$  varía en el rango  $\left[\frac{1}{3}, \frac{2}{3}\right]$

Sokal & Sneath\_4

$$A_{68} = \frac{a + d}{a + \frac{1}{2}(b + c) + d}$$

$$= 0 \text{ si } d = n$$

Yule\_1

$$A_{69} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

$$= 1 \text{ si } a = n \text{ ó } d = n$$

$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0 \text{ ó } (b + d) = 0 \text{ ó } (c + d) = 0$$

Sorensen

$$A_{70} = \frac{2a}{2a + b + c}$$
$$= 0 \text{ si } d = n$$

Sorgenfrei

$$A_{71} = \frac{a^2}{(a + b)(a + c)}$$
$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0$$

Sokal & Sneath\_5

$$A_{75} = \frac{ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$
$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0 \text{ ó } (b + d) = 0 \text{ ó } (c + d) = 0$$

Yule\_2

$$A_{76} = \frac{ad - bc}{ad + bc}$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$
$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0 \text{ ó } (b + d) = 0 \text{ ó } (c + d) = 0$$

Yule\_3

$$A_{77} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$
$$= 1 \text{ si } a = n \text{ ó } d = n$$
$$= 0 \text{ si } (a + b) = 0 \text{ ó } (a + c) = 0 \text{ ó } (b + d) = 0 \text{ ó } (c + d) = 0$$

## 2. 4. Propiedades de las matrices de similitud y de disimilitud

Los índices de similitud y/o disimilitud, como fue expresado en la Introducción de este trabajo, son utilizados como input en numerosas aplicaciones de métodos de análisis multivariados, tanto de ordenamiento como de clasificación. En esos contextos son utilizados en conjunto, reuniendo en un arreglo cuadrado y simétrico los índices para todos los pares de observaciones, conformando matrices que pueden requerir, dependiendo de los métodos a aplicar, ciertas

propiedades, referidas a sus características y a su posible representación en un espacio euclídeo. Este trabajo se centra en los métodos de ordenamiento y en especial en Análisis de Coordenadas Principales (PCoA) ya que para la clasificación no hay restricciones particulares para los índices. Este método tiene la ventaja sobre otros (Análisis de Componentes Principales y Análisis de Correspondencias) que se puede utilizar cualquier medida de similitud/disimilitud para construir la matriz de entrada. Si la matriz de disimilitud es definida positiva o semi-definida positiva (psd) entonces PCoA da una representación euclídea reducida exacta, donde las distancias entre los puntos son interpretadas como las disimilitudes entre los correspondientes objetos y los ángulos de los vectores como medidas de similitud entre variables (Camiz 1999). En este esquema, las representaciones reducidas en dimensión son efectivas para mostrar una síntesis de la información y describir el conjunto de datos. Cuando la matriz no es ni definida positiva, ni psd y se utiliza PCoA, ocurren autovalores negativos. Esto es considerado una seria desventaja para la interpretación de los autovalores en términos de inercia (variancia) explicada y por lo tanto la representación euclídea se vuelve crítica. En su lugar, una aproximación euclídea es siempre posible a través de diferentes técnicas. Una alternativa es utilizar Escalamiento Multidimensional No métrico (por sus siglas en inglés NMMDS *Non Metric Multidimensional Scaling*), sin embargo, este es un método que consume tiempo computacional, depende de la configuración inicial, a veces conduce a un óptimo local y se pierde la información métrica si esta existe. Otra solución es ignorar los autovalores negativos que suelen ser pequeños y aportan poca información (Pielou 1984). Este procedimiento, fue demostrado por Mardia (1978) que tiene propiedades óptimas. Otra solución (Lingoes 1971, Cailliez 1983) es adicionar una constante adecuada a todas las distancias. Camiz (1999) demostró que esto conduce a ejes sesgados en la misma cantidad y que una solución mejor es considerar a los autovalores negativos números imaginarios, como así también las coordenadas de los individuos (Gower 1985). De esta manera, la inercia total puede ser particionada en dos partes: una positiva y otra negativa. La positiva sobreestima la inercia verdadera, con una representación euclídea real, esto es, una aproximación euclídea de los individuos en el espacio expresado por los autovectores reales. La parte negativa corresponde a la parte sobreestimada y proporciona una representación imaginaria (sobre los

autovectores imaginarios) del sesgo introducido en la aproximación euclídea. La explicación geométrica es que en un espacio euclídeo, las figuras geométricas no euclidianas no pueden ser inmersas, entonces, para que la representación de los puntos sea posible, algunas distancias deben ser alargadas, consecuentemente la inercia original es aumentada y se vuelve sobreestimada. La inercia adicionada no existe en los datos. Los autovalores negativos y los ejes imaginarios son necesarios para balancear la inercia con su contribución negativa y ellos explican dónde y cuánta inercia extra fue adicionada para permitir una representación euclídea. De aquí la importancia de evaluar la euclideanidad de las matrices de disimilitud construidas con los índices de disimilitud (Camiz 1999).

### **Propiedad 15: Euclideanidad**

Ninguna disimilaridad ni distancia se puede usar directamente para ubicar puntos en un sistema de coordenadas ortogonales en el cual la distancia entre puntos se expresa en la forma euclidiana habitual, o sea como la raíz cuadrada de la suma de los cuadrados de las diferencias entre coordenadas ( $D_i(j,k) = \sqrt{\sum_{h=1}^n (X_{jh} - X_{kh})^2}$ ). Esto solo se puede conseguir si la disimilitud  $D_i$  es una distancia y cumple con la quinta condición:

5) la distancia  $D_i$  entre dos puntos es una distancia Euclideana si es la norma del vector que une los dos puntos, norma que resulta de un producto escalar definido positivo.

A un espacio con una distancia Euclídea se lo llama *espacio Euclídeo*.

Datos  $r$  vectores  $\mathbf{X}_j$  ( $j=1, \dots, r$ ), con  $\mathbf{D}_{rxr}$  matriz de distancias entre  $\mathbf{X}_j$  y  $\mathbf{X}_k$ ,  $D_i = D_i(j, k)$  (para  $j, k=1, \dots, r$ ), que cumple con las cinco condiciones, se puede efectuar su inmersión Euclideana, o sea representarlos en el espacio gráfico ordinario, donde la distancia se calcula como la longitud del vector de diferencia, calculada bajo el teorema de Pitágoras.

Una definición equivalente es la siguiente:

5)  $D_i$  es una distancia Euclídeana si la matriz de Torgerson:

$$S_{jk} = -\frac{1}{2}(D_{jk}^2 - D_{j.}^2 - D_{.k}^2 + D_{..})$$

$\forall j,k=1,\dots,r$  correspondiente a  $D_{rxr}$  es semi definida positiva.

En ciertas ocasiones, esta propiedad se puede comprobar algebraicamente, si no hay que verificar caso por caso si la matriz  $D_{rxr}$  tiene o no autovalores negativos.

Para una prueba de la misma vease Gower (1982).

**Teorema 1:** Una matriz  $D_{rxr}$  es Euclídea si y solo si la matriz  $S_{rxr}=(I-s\mathbf{1}')\Delta(I-s\mathbf{1}')$  es semi definida positiva (todos sus autovalores son no negativos), con  $s\mathbf{1}'=1$ , donde:

$I_{rxr}$  es la matriz identidad de orden  $r$

$\mathbf{1}_{rx1}$  es un vector columna de unos, de dimensión  $r$

$\Delta_{rxr}$  es una matriz cuadrada de orden  $r$  con elementos  $\{-1/2 D_i(j, k)^2\}$

$s_{rx1}$  es un vector columna de dimensión  $r$

Se demuestran los siguientes resultados: Si  $A_{rxr}$  es una matriz de similitud semi definida positiva (psd) con elementos  $A_i(j,k)$  variando entre 0 y 1 y  $A_i(j,j)=1$ , luego la matriz de disimilitud derivada de ella con elementos  $D_i(j, k) = \sqrt{1 - A_i(j, k)}$  es Euclídea (Gower & Legendre 1986) y también lo es  $D_i(j, k) = \sqrt{2(1 - A_i(j, k))}$  (Gower 1966). Estas son condiciones suficientes pero no necesarias, ya que podemos encontrar una matriz de similitud  $A_{rxr}$  que no sea psd y sin embargo, la matriz de disimilitud con elementos  $D_i(j, k)$  derivada de ella puede ser Euclídea.

## 2. 5. Diseño del estudio comparativo

Las propiedades enunciadas resultan disímiles en cuanto a la forma de comprobación que ellas requieren. Algunas pueden verificarse sencillamente observando el cumplimiento de las condiciones en forma aritmética. Esto es evidente para las propiedades de: Existencia de valores mínimos y máximos

(prop.3), Asociación completa o absoluta (prop.4) y Asociación nula cuando el valor de  $a=0$  (prop.5).

Para la verificación de otras propiedades como Existencia de límites convencionales (prop.1), Maximal y Minimal (prop.2), Simetría (prop.6), Monotonicidad respecto de la estadística  $\chi^2$  (prop.7), Distribuciones en el muestreo (prop.11), Métrica (prop.13), Indeterminancia (prop.14) y Euclideanidad (prop.15) debe recurrirse a un procedimiento de simulación. El proceso utilizado consistió en simular una matriz de datos binarios aleatoriamente considerando un número fijo de unidades de análisis y cambiando las probabilidades de los resultados de ausencia (0) y presencia (1). Se simularon 100 ( $r$ ) vectores de datos de dimensión diez ( $n$ ) en veinte secuencias que utilizan distintas semillas aleatorias. Cada secuencia de cinco vectores genera cada elemento a partir de una distribución Binomial puntual con diferentes valores para la probabilidad de suceso. La matriz de dimensión  $100 \times 10$  resultante ( $X_{\text{simulada}}$ ) se utiliza para la verificación de distintas propiedades. En estudios de Ecología vegetal, estos vectores representan unidades muestrales de vegetación donde cada elemento (0 ó 1) indica la presencia/ausencia de una especie vegetal con diferente probabilidad en la región o unidad observada. Las situaciones consideradas fueron las siguientes:

- 1) las diez especies tienen probabilidad de presencia creciente ( $p=0.05, 0.10, 0.25, 0.30, 0.45, 0.50, 0.65, 0.70, 0.85, 0.90$ ),
- 2) siete especies vegetales tienen probabilidad de presencia baja ( $p=0.05$  ó  $0.10$ ) y otras tres, probabilidad alta ( $p=0.70, 0.85, 0.90$ ),
- 3) siete especies tienen probabilidad de presencia alta ( $p=0.85$  ó  $0.90$ ) y otras tres, probabilidad baja ( $p=0.05, 0.10, 0.25$ ),
- 4) cinco especies tienen probabilidad baja ( $p=0.25$ ) y cinco probabilidad alta ( $p=0.60$ ),
- 5) cinco especies tienen probabilidad muy baja ( $p=0.05$ ) y cinco muy alta de suceso ( $p=0.85$ ).

Para probar la propiedad de Linealidad (prop.8) se utilizaron los escenarios definidos por Hubálek (1982), fijando un valor  $n=100$  y cuatro variantes de  $a$ ,  $b$ ,

*c* y *d*, manteniendo los mismos totales marginales (Tabla 2) y con un criterio más general, los escenarios de Gower & Legendre (1986) denominados OCCAS (Tabla 3) y que contemplan veintisiete variantes de las cuatro cantidades de frecuencias que están igualmente espaciadas con totales marginales fijos y variables.

*Tabla 2: Escenarios para estudiar la propiedad de linealidad de índices de similitud para datos binarios.*

		Obs. $x_k$					Obs. $x_k$		
		presente	ausente				presente	ausente	
Obs. $x_i$	presente	0	40	40	Obs. $x_i$	presente	1	39	40
	ausente	40	20	60		ausente	39	21	60
		40	60	n=100			40	60	n=100

		Obs. $x_k$					Obs. $x_k$		
		presente	ausente				presente	ausente	
Obs. $x_i$	presente	2	38	40	Obs. $x_i$	presente	3	37	40
	ausente	38	22	60		ausente	37	23	60
		40	60	n=100			40	60	n=100

Las variantes de la Tabla 3 permiten obtener los indicadores definidos como resolución, no linealidad y CV.

Tabla 3: Escenarios para estudiar la propiedad de linealidad de índices de similitud para datos binarios.

OCCAS 1 (9 valores)		Obs. $x_k$			OCCAS 2 (9 valores)		Obs. $x_k$		
		presente	ausente				presente	ausente	
Obs. $x_i$	presente	45->5	5->45	50	Obs. $x_i$	presente	90->10	10->90	100
	ausente	50	50	100		ausente	0	50	50
		95->55	55->95	n=150			90->10	60->140	n=150
OCCAS 3 (9 valores)		Obs. $x_k$			OCCAS 1: $a$ disminuye de la misma manera en que $b$ aumenta, en un valor de 5, manteniendo $a+b$ constante en un valor de 50, por su parte $c$ y $d$ tienen valor constante. OCCAS 2: sigue el mismo esquema que la anterior, pero los cambios en $a$ y $b$ son en un valor de 10, mientras que $c=0$ y $d$ es constante. OCCAS 3: $a$ disminuye en un valor de 10, mientras que $b$ y $c$ aumentan en un valor de 5 cada uno, manteniendo $a+b+c$ constante en un valor de 100 y $d$ es constante.				
		presente	ausente						
Obs. $x_i$	presente	90->10	5->45	95->55					
	ausente	5->45	50	55->95					
		95->55	55->95	n=150					

Para probar las propiedades de Insensibilidad al valor de  $a=0$  (prop.9) y asociación completa o absoluta (prop.10) se utiliza el escenario de Hubálek utilizado para la propiedad de Linealidad (prop.8) ( $n=100$ ,  $a=0$ ,  $b=40$ ,  $c=40$  y  $d=20$ ) y se completa con todas las iteraciones posibles ( $a+1$ ,  $b-1$ ,  $c-1$ ,  $d+1$ ) hasta llegar a los siguientes valores:  $a=40$ ,  $b=0$ ,  $c=0$  y  $d=60$ , manteniendo siempre los mismos totales marginales y  $n=100$  (Tabla 4). En cada una de las iteraciones se calculan todos los índices estudiados en este trabajo.

Tabla 4: Escenarios definidos para estudiar la propiedad de Insensibilidad al valor de  $a=0$  y Asociación completa o absoluta de índices de similitud para datos binarios.

		Obs. $x_k$			→			Obs. $x_k$		
		presente	ausente					presente	ausente	
Obs. $x_i$	presente	0	40	40		presente	40	0	40	
	ausente	40	20	60		ausente	0	60	60	
		40	60	n=100			40	60	n=100	

Para verificar la propiedad de Insesgamiento (prop.12) se consideran cinco escenarios con distintos valores conocidos y constantes para las frecuencias de celdas:  $a$ ,  $b$ ,  $c$  y  $d$  como valores poblacionales (Tabla 5). Estos escenarios coinciden con los propuestos por Goodall en 1973.

Tabla 5: Escenarios para estudiar la propiedad de insesgamiento de índices de similitud para datos binarios.

		Obs. $x_k$						Obs. $x_k$			
		presente	ausente					presente	ausente		
Obs. $x_i$	presente	1	29	30	Obs. $x_i$	presente	2	3	5		
	ausente	30	40	70		ausente	45	50	95		
		31	69	$n=100$			47	53	$n=100$		
		Obs. $x_k$					Obs. $x_k$				
		presente	ausente				presente	ausente			
Obs. $x_i$	presente	5	20	25	Obs. $x_i$	presente	10	2	12		
	ausente	35	40	75		ausente	18	70	88		
		40	60	$n=100$			28	72	$n=100$		
		Obs. $x_k$					Obs. $x_k$				
		presente	ausente				presente	ausente			
Obs. $x_i$	presente	30	10	40							
	ausente	10	50	60							
		40	60	$n=100$							

Los diferentes escenarios consideran distintos valores para  $a$ ,  $b$ ,  $c$  y  $d$  y por lo tanto también cambian los totales marginales, no así el total que permanece constante ( $n=100$ )

Para cada caso definido se calculan los índices de similitud, los que son considerados valores poblacionales. Luego se generan, para cada escenario, 1000 muestras de diferentes tamaños: 10, 20, 50 y 100 de una distribución multinomial con valores fijos de probabilidad de cada celda definida por las frecuencias  $a$ ,  $b$ ,  $c$  y  $d$ . Se calculan los índices de similitud, para cada repetición, escenario y tamaño de muestra, calculando cobertura, sesgos y SE por cada tamaño de muestra.

Para probar la propiedad Métrica (prop.13) sobre las matrices de disimilitud una vez verificados el signo, el valor cero de distancia consigo mismo y la simetría,

se consideran para cada índice todas las posibles variantes de ternas entre los elementos generados en la matriz simulada  $X_{\text{simulada}}$  y para cada terna se verifica si se cumple la desigualdad triangular entre las distancias.

Para probar la propiedad de Euclideanidad (15) se fue testando si la matriz de Torgerson (1958) correspondiente a cada índice de disimilitud tiene autovalores negativos.

Todo el procedimiento realizado para el estudio de propiedades fue sistematizado e implementado computacionalmente utilizando software libre R y SAS disponible en los laboratorios de la Escuela de Estadística (Fac. de Cs. Económicas y Estadística, U.N.R.) por convenio con la compañía.

Para verificar los ajustes de las distribuciones empíricas a la distribución normal (prop.11) se utilizó el paquete `fitdistrplus` (Delignette-Muller et al. 2013) del software R. Se emplea además un procedimiento bootstrap (Efron & Tibshirani 1994).

Para algunas propiedades se construyeron programas en lenguaje IML de SAS (versión 9). En el Anexo 8 se encuentran para cada propiedad el o los programas tipo, en particular correspondiente al índice de similitud de Jaccard. El listado de los programas tipo es:

Propiedad 2: Maximal y Minimal → Programa 107 (página 299)

Propiedad 6: Simetría → Programa 108 (página 303)

Propiedad 7: Monotonicidad respecto de la estadística  $\chi^2$  → Programas 106 y 99 (páginas 307 y 309 respectivamente)

Propiedad 8: Linealidad → Programas 200 y 410 (páginas 311)

Propiedad 9: Insensibilidad al valor de  $a=0$  → Programa 100 (página 315)

Propiedad 11: Distribuciones en el muestreo → Programa 104 (página 316)

Propiedad 12: Insegamiento → Programa 211 (página 319)

Propiedad 13: Métrica → Programa 758 (página 321)

Propiedad 15: Euclideanidad → Programa 662 (página 325)

Todas las figuras fueron elaboradas con el paquete `ggplot` de R.

## CAPITULO 3

### RESULTADOS Y DISCUSIÓN

Se presentan en este capítulo los resultados encontrados acerca del cumplimiento de las quince propiedades estudiadas para los setenta y siete índices de similitud seleccionados y las tres disimilitudes obtenidas a partir de ellos. Para las similitudes se prueban catorce propiedades, todas menos la métrica (prop.13), que es exclusiva de las disimilitudes. Para cada disimilitud se prueban seis propiedades: Maximal y Minimal (prop.2), Simetría (prop.6), Distribuciones en el muestreo (prop.11), Métrica (prop.13), Indeterminancia (prop.14) y Euclídeanidad (prop.15). A fin de resumir los hallazgos sobre cada propiedad y facilitar su análisis se presentan tablas y/o figuras con los resultados encontrados. Finalmente, a modo de síntesis comparativa se presenta una tabla que reúne simultáneamente a todas las propiedades para el conjunto total de índices de similitud y tres tablas adicionales, para cada una de las disimilitudes estudiadas.

#### **Propiedad 1: Existencia de límites convencionales**

En general, los autores que plantean las distintas propuestas de índices de similitud o disimilitud, analizan sus valores mínimos y máximos como particularidades. Sin embargo, en variadas situaciones este análisis no se incluye. Esta situación se encontró en 34 de los 77 índices estudiados (44%).

En este apartado se registran y comparan los valores extremos de todos los índices de similitud, teniendo en cuenta los valores publicados, en caso de haber sido éstos reportados en la literatura. Para algunos casos es necesario analizar las distribuciones en el muestreo obtenidas a partir de simulación para derivar sus valores mínimos y máximos.

Un alto porcentaje de los índices (70%) caen dentro de los dos rangos definidos por Hubálek (1982) y considerados como convencionales, es decir:

$$\text{caso 1: } -1 \leq A_i \leq 1$$

$$\text{caso 2: } 0 \leq A_i \leq 1$$

Estas dos categorías fueron identificadas por Dagnelie en 1960 como índices centrados (caso 1) y no centrados (caso 2).

El caso que se presenta más frecuentemente corresponde a los índices que varían en el intervalo  $[0,1]$  (56%). En un orden menor (44%) aparece el grupo de aquellos con rango de variación  $[-1,1]$ .

Dentro de los índices que no tienen límites convencionales se observa una gran variedad de rangos. En primer lugar se tienen cinco índices (Forbes\_1, Fossum, Johnson\_2, Kulczynski\_1 y Sokal & Sneath\_3) que varían entre 0 y  $+\infty$ , que Hubalek (1982) también considera límite convencional, pero no en este trabajo. Otros cinco índices (Anónimo\_5, Dennis, Forbes\_3, Gilbert & Wells y Stiles) también tienen límites no finitos y por lo tanto no permiten definir el 100% de similitud y resultan los más controvertidos. Además, al no estar acotados no es posible calcular a partir de ellos las disimilitudes asociadas.

En la Tabla 6 se puede observar el listado de los índices que tienen alguno de estos límites convencionales (casos 1 y 2) y también aquellos que no (30%). Para facilitar las comparaciones se ha organizado la Tabla agrupando los índices según pertenezcan o no a las dos categorías definidas (casos 1 y 2). El listado deja las posiciones finales a los índices con límites no convencionales. Estos se dividen en aquellos acotados, que permiten derivar las disimilitudes y los que no son acotados. Dentro de cada categoría los índices se ordenan alfabéticamente por autor para facilitar su búsqueda.

Los índices más utilizados y más comúnmente aconsejados por la mayoría de los autores son los que tienen límites finitos, que además permiten pasar muy fácilmente de una similitud a una disimilitud.

El caso de rango superior infinito, es desaconsejado ya que la similitud total no puede ser definida y estos coeficientes, en general, son sensibles a pequeños cambios en el valor de  $a$  (Clifford & Stephenson 1975).

De los cuarenta y tres índices estudiados por Hubalek (1982) no tienen límites convencionales Johnson\_1, Eyraud y Gilbert & Wells, lo cual coincide con lo

obtenido. Sin embargo Fager & Mc Gowan y Goodall no los tienen porque hay errores en las fórmulas de los mismos que figuran en su trabajo.

En todos los casos en que los valores extremos no fueran fácilmente identificables por su definición, se realizaron procedimientos de simulación a fin de identificarlos. En algunos casos se corroboraron los límites explicitados en la literatura, en otros se constataron diferencias respecto de los valores publicados y para los índices con límites desconocidos, se encontraron sus valores.

Los índices de Forbes\_2 y Eyraud no alcanzan el límite inferior, los índices de Tarwid, Faith, Pearson\_2 y Russell & Rao no alcanzan el límite superior, mientras que Fager & Mc Gowan no alcanza ninguno de los dos.

Hubálek (1982) encontró los mismos resultados para Tarwid y Pearson\_2, también indicó que el índice de Cole\_1 no alcanza el límite superior mientras en este trabajo sí lo alcanza, sin embargo la fórmula del mismo tiene un error (signo menos en el denominador en lugar de tener un signo más).

Tabla 6: Rangos de variación de los índices de similitud para datos binarios.

Límites	Indice de similitud	Rango de variación	Mínimos y Máximos (*)		
convencionales	Anónimo_1				
	Anónimo_3				
	Baroni-Urbani & Buser_2				
	Cohen				
	Cole_1				
	Digby				
	<b>Fager &amp; Mc Gowan</b>				
	<b>Forbes_2</b>				
	Goodman & Kruskal				
	Hamann				
	Hurlbert_1				
	Caso 1	Hurlbert_2	$-1 \leq A_i \leq 1$		
		Kendall			
		Maxwell & Pilliner			
		Mc Connaughey			
		Michael			
		Pearson & Heron			
		Peirce_2			
		Peirce_3			
		Scott			
	<b>Tarwid</b>				
	Yule_1				
	Yule_2				
	Yule_3				
convencionales	Anónimo_2				
	Anónimo_4				
	Baroni-Urbani & Buser_1				
	Braun-Blanquet				
	Clifford & Stephenson				
	Cole_2				
	Dice_1				
	Dice_2				
	Doolittle				
	<b>Faith</b>				
	Goodall				
	Hawkins & Dotson				
	Jaccard				
	Kulczynski_2				
	Legendre & Legendre				
	Caso 2	Ochiai	$0 \leq A_i \leq 1$		
		<b>Pearson_2</b>			
		Peirce_1			
		Rogers & Tanimoto			
		Rogot & Goldberg			
	<b>Russell &amp; Rao</b>				
	Simpson				
	Sokal & Michener				
	Sokal & Sneath_1				
	Sokal & Sneath_2				
	Sokal & Sneath_4				
	Sokal & Sneath_5				
	Sorensen				
	Sorgenfrei				
	Tschuproff				
NO convencionales	acotados	Clement	$[0, n]$		
		Cole_3	$[-\infty, 1]$		
		Cole_4	$[-\infty, 1]$		
		Dice_3	$[0, 2]$		
		<b>Eyraud</b>	$[-1, 0]$		
		Fleiss	$[-\infty, 1]$		
		Harris & Lahey	$[0, n]$		
		Johnson_1	$[0, 2]$		
		Kuder & Richardson	$[-\infty, 1]$		
		Lamont & Grant	$[0, 0.5]$		
		Loevinger	$[-\infty, 1]$		
		Mountford	$[0, 2]$		
		Pearson_1	$[0, n]$		
		NO acotados	Anónimo_5	$[-1, +\infty]$	
		Dennis	$[-\infty, +\infty]$		
	Forbes_1	$[0, +\infty]$			
	Forbes_3	$[-1, +\infty]$			
	Fossum	$[0, +\infty]$			
	Gilbert & Wells	$[-\infty, +\infty]$			
	Johnson_2	$[0, +\infty]$			
	Kulczynski_1	$[0, +\infty]$			
	Sokal & Sneath_3	$[0, +\infty]$			
	Stiles	$[-\infty, +\infty]$			

(\*) los mínimos y máximos solo se reportan cuando no coinciden con los extremos del rango de variación

En la Sección 2. 2. de Definiciones se destaca la relación entre los valores extremos de la similitud, cuando estos son finitos ([0,1] o [-1,1]) y de las disimilitudes derivadas de la anterior.

Cuando se pasa de una similitud a una disimilitud, si la similitud ( $A_i$ ) varía entre 0 y 1, la disimilitud  $D1_i = 1 - A_i$  tiene el mismo rango de variación. Para la disimilitud  $D2_i = \sqrt{1 - A_i}$  el rango de variación sigue siendo [0,1], pero para  $D3_i = \sqrt{2(1 - A_i)}$  el rango de variación cambia y es [0,  $\sqrt{2}$ ].

De la misma manera para las similitudes  $A_i$  en el rango -1 y 1, la disimilitud  $D1_i = \frac{1-A_i}{2}$  varía en el rango [0,1], la disimilitud  $D2_i = \sqrt{\frac{1-A_i}{2}}$  también varía entre 0 y 1, mientras que  $D3_i = \sqrt{2(\frac{1-A_i}{2})} = \sqrt{1 - A_i}$  varía en el rango [0,  $\sqrt{2}$ ].

Algunos de los índices tratados en este trabajo no tienen límites convencionales, por lo que las disimilitudes derivadas de ellos deben ser redefinidas para que resulten con rangos casi convencionales. La Tabla 7 resume estas relaciones para los casos no convencionales acotados.

Un pequeño porcentaje de índices de similitud (13%) no son acotados (dentro de los no convencionales), no permitiendo por lo tanto el cálculo de las disimilitudes. Los índices son: Anónimo\_5, Dennis, Forbes\_1, Forbes\_3, Fossum, Gilbert & Wells, Johnson\_2, Kulczynski\_1, Sokal & Sneath\_3 y Stiles.

Tabla 7: Nuevos rangos de variación para las disimilitudes obtenidas de índices de similitud con límites no convencionales.

Índice de similitud $A_i$	Similitud $A_i$	Disimilitud $D1_i$	Disimilitud $D2_i$	Disimilitud $D3_i$
		$D1_i=n-A_i$	$D2_i=\sqrt{(n-A_i)}$	$D3_i=\sqrt{2(n-A_i)}$
Clement				
Harris & Lahey	$[0,n]$	$[0,n]$	$[0,\sqrt{n}]$	$[0,\sqrt{2n}]$
Pearson_1				
		$D1_i=1-A_i$	$D2_i=\sqrt{(1-A_i)}$	$D3_i=\sqrt{2(1-A_i)}$
Cole_3				
Cole_4				
Fleiss	$[-\infty,1]$	$[0,+\infty]$	$[0,+\infty]$	$[0,+\infty]$
Kuder & Richardson				
Loevinger				
		$D1_i=2-A_i$	$D2_i=\sqrt{(2-A_i)}$	$D3_i=\sqrt{2(2-A_i)}$
Dice_3				
Johnson_1	$[0,2]$	$[0,2]$	$[0,\sqrt{2}]$	$[0,2]$
Mountford				
		$D1_i=1+A_i$	$D2_i=\sqrt{(1+A_i)}$	$D3_i=\sqrt{2(1+A_i)}$
Eyraud	$[-1,0]$	$[0,1]$	$[0,1]$	$[0,\sqrt{2}]$
		$D1_i=0.5-A_i$	$D2_i=\sqrt{(0.5-A_i)}$	$D3_i=\sqrt{2(0.5-A_i)}$
Lamont & Grant	$[0,0.5]$	$[0,0.5]$	$[0,\sqrt{0.5}]$	$[0,1]$

## Propiedad 2: Maximal y Minimal

De los setenta y siete índices de similitud estudiados, el 90% cumplen con esta propiedad, es decir, alcanzan el valor máximo cuando se calcula la similitud de un vector consigo mismo. Solamente dos índices no la cumplen: Eyraud y Michael (3%). Cinco índices (7%) resultan indeterminados cuando se calcula la similitud de un vector consigo mismo, ya que en esta situación se anulan las expresiones del denominador. Son los índices de: Johnson\_2, Kulczynski\_1, Mountford, Peirce\_1 y Sokal & Sneath\_3.

Los mismos resultados se obtienen para las tres disimilitudes derivadas de las similitudes, pero en estos casos se estudia si las mismas alcanzan el valor mínimo cuando se calcula la disimilitud de un vector consigo mismo.

Para siete índices (9%) no acotados superiormente, no se pudieron calcular las disimilitudes, por lo tanto los resultados se limitan a las similitudes, observando que todos la cumplen, Anónimo\_5, Dennis, Forbes\_1, Forbes\_3, Fossum, Gilbert & Wells y Stiles.

### **Propiedad 3: Valores mínimos y máximos**

Esta propiedad se divide en dos partes. Una de ellas se denomina 3a) y se refiere a si los índices de similitud alcanzan el valor mínimo cuando  $a=d=0$ . El 74% de los índices considerados (77) cumplen con la propiedad, mientras que el 18% no la satisfacen, ellos son: Anónimo\_4, Clifford & Stephenson, Cole\_2, Dennis, Doolittle, Fager & Mc Gowan, Fleiss, Fossum, Loevinger, Pearson\_1, Pearson\_2, Peirce\_1, Stiles y Tschuproff. Por su parte seis índices (8%) del total, solo cumplen la propiedad bajo ciertos requisitos adicionales. Por ejemplo, el índice de Eyraud la cumple solo si  $bc=1$ . Esta condición, por tratarse de números enteros, implica que  $b=c=1$ . Por su parte los índices de Cohen, Cole\_3, Cole\_4, Forbes\_2 y Michael solo la cumplen si  $b=c$ . Las demostraciones se encuentran en el Anexo 1.

Respecto de la parte b) de la Propiedad 3, esta es satisfecha por 60 de 77 índices estudiados (78%), es decir, que una alta proporción alcanzan el valor máximo cuando  $b=c=0$ . Los índices que no la cumplen son: Anónimo\_5, Dennis, Eyraud, Fager & Mc Gowan, Faith, Forbes\_1, Forbes\_3, Fossum, Gilbert & Wells, Mountford, Pearson\_2, Peirce\_1, Russell & Rao, Stiles y Tarwid (19%). Por su parte el porcentaje restante (3%) solo la cumple bajo ciertos requisitos adicionales, el índice de Dice\_3 solo si  $a=1$  y el índice de Michael solo si  $a=d$ . Las demostraciones se encuentran en el Anexo 2.

Más del 80% de los índices que cumplen la propiedad 3a) también cumplen la 3b). Una pequeña proporción de índices son coincidentes en la falta de cumplimiento de ambas: Dennis, Fager & Mc Gowan, Fossum, Pearson\_2, Peirce\_1 y Stiles.

Los resultados obtenidos coinciden con lo hallado por Hubalek (1982) en el grupo por él estudiado, quien encontró incompatibles con esta propiedad a los índices de Doolittle, Fager & Mc Gowan, Pearson\_1, Pearson\_2 y Peirce\_1, mientras que el índice de Michael la cumplía solo si  $b=c$  para la parte a) y si  $a=d$  para la parte b). También encuentra que el índice de Cole\_1 no la cumple, sin embargo como se menciona en la prop.1, la fórmula del mismo tiene un error.

El comportamiento en lo que se refiere a esta propiedad, partes a) y b) se resume en la Tabla 8.

#### **Propiedad 4: Asociación completa o absoluta**

Para probar esta propiedad es necesario que los índices cumplan la Propiedad 3b), es decir que alcancen el valor máximo cuando  $b=c=0$ . De esta forma se puede probar si también alcanzan el valor máximo cuando solamente  $b=0$  o  $c=0$ , alcanzando la asociación completa, en caso contrario se dice que la asociación es absoluta. El 21% de los índices no cumplen la Propiedad 3b) y no permiten probar la Propiedad 4. Un alto porcentaje de los índices (82%) cumplen la condición de asociación absoluta, mientras que solo seis índices (10%) cumplen la asociación completa, estos son: Digby, Loevinger, Pearson & Heron, Simpson, Yule\_2 y Yule\_3 (Tabla 8).

Un conjunto de cinco índices (8%) tienen un comportamiento diferenciado según que  $b=0$  o  $c=0$ . Los índices de Cole\_3, Dice\_1 y Johnson\_2, cuando  $b=0$  coinciden con el valor del índice para  $b=c=0$ , que es el valor máximo, alcanzando la condición de asociación completa, mientras que para  $c=0$  toman un valor menor que el máximo, comportándose como un índice de asociación absoluta. Los índices de Cole\_4 y Dice\_2 tienen el comportamiento inverso, cuando  $c=0$  coinciden con el valor del índice para  $b=c=0$ , que es el valor máximo, alcanzando la condición de asociación completa, mientras que para  $b=0$  toman un valor menor que el máximo, comportándose como un índice de asociación absoluta. Por su parte el índice de Dice\_3, que cumple la Propiedad 3b) solamente si  $a=1$ , y el índice de Michael que la cumple solo si  $a=d$ , bajo esas situaciones, ambos se comportan como índices de asociación absoluta. Las demostraciones se encuentran en el Anexo 3.

Hubálek (1982) también encontró que la mayoría de los índices estudiados por él satisfacen la propiedad de asociación absoluta, mientras que unos pocos la asociación completa: Forbes\_2, Hurlbert\_1, Hurlbert\_2, Peirce\_1, Pearson & Heron, Simpson, Yule\_2 y Yule\_3. En este trabajo no hemos encontrado los mismos resultados para los cuatro primeros índices, es más, Peirce\_1 no cumple la propiedad 3b) por lo tanto no permite calcular esta propiedad.

#### **Propiedad 5: Asociación nula**

Los índices que cumplen esta propiedad son menos numerosos (43%) que aquellos que cumplen la parte a) de la Propiedad 3 (74%), ya que esta propiedad

es más estricta al exigir que un índice alcance el menor valor cuando solamente  $a=0$ . Todos los índices que cumplen la Propiedad 5 cumplen además la 3a), excepto el índice Anónimo\_4, que da como resultado  $A_4(\mathbf{j}, \mathbf{k})^{(a=0,b,c,d=0)}=0/0$  siendo indeterminado para la Propiedad 3a), mientras que  $A_4(\mathbf{j}, \mathbf{k})^{(a=0,b,c,d=0)}=0$  (mínimo) para la Propiedad 5 (Tabla 8). Las demostraciones se encuentran en el Anexo 4.

Los coeficientes de Eyraud, Goodall, Hamann, Michael, Peirce\_2, Roger & Tanimoto, Sokal & Michener, Sokal & Sneath\_2, Sokal & Sneath\_3, Sokal & Sneath\_4 y Yule\_1 fueron reportados por Hubálek (1982) como incompatibles con esta condición y coinciden con los resultados hallados en este trabajo.

Se indican en la última columna de la Tabla 8 los índices que cumplen las Propiedades 3a), 3b), 4 y 5, que representan el 31% de los mismos. Para la Propiedad 4 no se discrimina entre asociación absoluta o completa.

Tabla 8: Propiedades de Existencia de valores mínimos y máximos, Asociación nula cuando el valor de  $a=0$  y Asociación completa o absoluta.

Indice de similitud	Existencia de valores mínimos	Asociación nula	Existencia de valores máximos	Asociación completa o absoluta	Cumplimiento simultáneo
Anónimo_1	mínimo	mínimo	máximo	absoluta	*
Anónimo_2	mínimo	NO	máximo	absoluta	
Anónimo_3	mínimo	NO	máximo	absoluta	
Anónimo_4	NO	mínimo	máximo	absoluta	
Anónimo_5	mínimo	NO	NO	-----	
Baroni-Urbani & Buser_1	mínimo	mínimo	máximo	absoluta	*
Baroni-Urbani & Buser_2	mínimo	mínimo	máximo	absoluta	*
Braun-Blanquet	mínimo	mínimo	máximo	absoluta	*
Clement	mínimo	NO	máximo	absoluta	
Clifford & Stephenson	NO	NO	máximo	absoluta	
Cohen	NO/mínimo si $b=c$	NO	máximo	absoluta	
Cole_1	mínimo	NO	máximo	absoluta	
Cole_2	NO	NO	máximo	absoluta	
Cole_3	NO/mínimo si $b=c$	NO	máximo	completa si $b=0$ absoluta si $c=0$	
Cole_4	NO/mínimo si $b=c$	NO	máximo	completa si $c=0$ absoluta si $b=0$	
Dennis	NO	NO	NO	-----	
Dice_1	mínimo	mínimo	máximo	completa si $b=0$ absoluta si $c=0$	
Dice_2	mínimo	mínimo	máximo	completa si $c=0$ absoluta si $b=0$	*
Dice_3	mínimo	mínimo	NO/máximo si $a=1$	absoluta si $a=1$	
Digby	mínimo	mínimo	máximo	completa	*
Doolittle	NO	NO	máximo	absoluta	
Eyraud	NO/mínimo si $bc=1$	NO	NO	-----	
Fager & Mc Gowan	NO	NO	NO	-----	
Faith	mínimo	NO	NO	-----	
Fleiss	NO	NO	máximo	absoluta	
Forbes_1	mínimo	mínimo	NO	-----	
Forbes_2	NO/mínimo si $b=c$	NO	máximo	absoluta	
Forbes_3	mínimo	mínimo	NO	-----	
Fossum	NO	NO	NO	-----	
Gilbert & Wells	mínimo	mínimo	NO	-----	
Goodall	mínimo	NO	máximo	absoluta	
Goodman & Kruskal	mínimo	mínimo	máximo	absoluta	*
Hamann	mínimo	NO	máximo	absoluta	
Harris & Lahey	mínimo	NO	máximo	absoluta	
Hawkins & Dotson	mínimo	NO	máximo	absoluta	
Hurlbert_1	mínimo	NO	máximo	absoluta	
Hurlbert_2	mínimo	NO	máximo	absoluta	
Jaccard	mínimo	mínimo	máximo	absoluta	*
Johnson_1	mínimo	mínimo	máximo	absoluta	*
Johnson_2	mínimo	mínimo	máximo	completa si $b=0$ absoluta si $c=0$	*
Kendall	mínimo	NO	máximo	absoluta	
Kuder & Richardson	mínimo	NO	máximo	absoluta	
Kulczynski_1	mínimo	mínimo	máximo	absoluta	*
Kulczynski_2	mínimo	mínimo	máximo	absoluta	*
Lamont & Grant	mínimo	mínimo	máximo	absoluta	*
Legendre & Legendre	mínimo	mínimo	máximo	absoluta	*
Loevinger	NO	NO	máximo	completa	
Maxwell & Pilliner	mínimo	NO	máximo	absoluta	
Mc Connaughey	mínimo	mínimo	máximo	absoluta	*
Michael	NO/mínimo si $b=c$	NO	NO/máximo si $a=d$	absoluta si $a=d$	
Mountford	mínimo	mínimo	NO	-----	
Ochiai	mínimo	mínimo	máximo	absoluta	*
Pearson & Heron	mínimo	mínimo	máximo	completa	*
Pearson_1	NO	NO	máximo	absoluta	
Pearson_2	NO	NO	NO	-----	
Peirce_1	NO	NO	NO	-----	
Peirce_2	mínimo	NO	máximo	absoluta	
Peirce_3	mínimo	NO	máximo	absoluta	
Rogers & Tanimoto	mínimo	NO	máximo	absoluta	
Rogot & Goldberg	mínimo	NO	máximo	absoluta	
Russell & Rao	mínimo	mínimo	NO	-----	
Scott	mínimo	NO	máximo	absoluta	
Simpson	mínimo	mínimo	máximo	completa	*
Sokal & Michener	mínimo	NO	máximo	absoluta	
Sokal & Sneath_1	mínimo	mínimo	máximo	absoluta	*
Sokal & Sneath_2	mínimo	NO	máximo	absoluta	
Sokal & Sneath_3	mínimo	NO	máximo	absoluta	
Sokal & Sneath_4	mínimo	NO	máximo	absoluta	
Sokal & Sneath_5	mínimo	mínimo	máximo	absoluta	*
Sorensen	mínimo	mínimo	máximo	absoluta	*
Sorgenfrei	mínimo	mínimo	máximo	absoluta	*
Stiles	NO	NO	NO	-----	
Tarwid	mínimo	mínimo	NO	-----	
Tschuproff	NO	NO	máximo	absoluta	
Yule_1	mínimo	NO	máximo	absoluta	
Yule_2	mínimo	mínimo	máximo	completa	*
Yule_3	mínimo	mínimo	máximo	completa	*

Nota: la última columna corresponde al cumplimiento simultáneo de las propiedades

### **Propiedad 6: Simetría**

El 77% del total de índices de similitud (77) cumplen con la propiedad de simetría. Las disimilitudes derivadas de índices simétricos también lo son. Los índices que no cumplen esta propiedad son: Clement, Cole\_3, Cole\_4, Dice\_1, Dice\_2, Peirce\_1, Peirce\_2 y Peirce\_3 (10%) (similitud y disimilitud). En el subgrupo de diez índices de similitudes no acotados solo uno de ellos no la cumple (Johnson\_2).

Con respecto a esta propiedad se han encontrado desacuerdos en la bibliografía respecto a los índices de Braun-Blanquet, Fager & Mc Gowan y Simpson. Estos, como se desprende de su definición, son simétricos, es decir, no se alteran al intercambiar las cantidades  $b$  y  $c$ , sin embargo Hubálek (1982) y Giansanti (1990) los consideran asimétricos, mientras que otros autores (Bren & Batagelj 2006, Warrens 2008, Demey et al. 2011) también consideran simétricos a Braun-Blanquet y Simpson como en este trabajo. Por su parte el índice de Fager & Mc Gowan tal como está definido erróneamente en Hubálek (1982) sería asimétrico.

### **Propiedad 7: Monotonicidad respecto de la estadística $\chi^2$**

Para probar esta propiedad se consideran todos los pares posibles de unidades en el conjunto de datos simulados descrito en la Sección 2. 5. (página 49) y para cada uno de los pares se calcula el valor de la estadística  $\chi^2$  y los valores de los índices  $A_i$ . El ordenamiento de  $\chi^2$  debería ser similar al logrado con cada  $A_i$  ( $i=1, \dots, 77$ ) para que la monotonicidad se satisfaga. Esta concordancia se mide a través de los coeficientes de correlación de rangos de Spearman y de Kendall. Se eligen los coeficientes no paramétricos por considerarlos medidas más pertinentes a la evaluación que se pretende, ya que no se busca evaluar la existencia de una relación lineal entre ambas medidas ( $\chi^2$  y  $A_i$ ) sino comprobar la concordancia entre las magnitudes de esos valores.

Casi la totalidad de los índices estudiados (97%) satisfacen la propiedad de monotonicidad, mostrando una correlación positiva significativa ( $p \leq 0,0001$ ) y

en general alta, superior a 0.5 (91%), considerando los dos coeficientes de correlación: Spearman y Kendall (Tabla 9).

Los dos únicos índices que no muestran correlación positiva estadísticamente significativa son Eyraud y Peirce\_1, incluso las correlaciones encontradas son negativas y/o casi cero. Los valores de p para Eyraud son  $p=0,4590$  y  $p=0,7336$  y para Peirce  $p=0,0177$  y  $p=0,0250$  correspondiendo a Spearman y Kendall respectivamente.

La Tabla 9 presenta el listado de las correlaciones entre  $\chi^2$  y  $A_i$  medidos según las dos variantes de coeficientes considerados: Spearman y Kendall.

Los índices se agrupan considerando las tres categorías de magnitud expuestas en el análisis de estos resultados, a saber: correlaciones fuertes con  $0,90 \leq r \leq 1$ , correlaciones moderadas con  $0,50 \leq r < 0,90$  y correlaciones débiles con  $r < 0,50$  con respecto al coeficiente de correlación de rangos de Spearman.

Hubálek (1982) probó la propiedad de Monotonicidad en algunos coeficientes de asociación utilizando datos reales obtenidos durante un estudio de campo y de laboratorio sobre la ocurrencia de una especie de hongo del género *Chaetomium Kunze ex Fries (Ascomycetes)* en 869 muestras tomadas de los nidos de pájaros viviendo en libertad (Hubálek 1974). Se encontraron siete especies distintas que contenían chaetomia. La particularidad de estos datos es que el valor de las ausencias conjuntas  $d$  es muy superior a los demás valores, incluso  $d > a+b+c$ .

Hubálek (1982) probó la monotonicidad con respecto a  $\sqrt{\chi^2}$  en treinta y dos índices utilizando el coeficiente de correlación momento-producto de Pearson. Resultaron inadmisibles los coeficientes de Goodall, Hamann, Roger & Tanimoto, Sokal & Michener, Sokal & Sneath\_3 y Sokal & Sneath\_4. El estudio realizado en esta Tesis muestra que estos índices son admisibles con correlaciones moderadas,  $0,6191 \leq r \leq 0,6340$  destacando el índice de Hamann cuya correlación con  $\sqrt{\chi^2}$  alcanza el valor:  $r=0,9091$ . Estos índices son los únicos, dentro de los considerados por Hubálek (1982), que tienen en el numerador la expresión  $a+d$ , por eso resulta probable que con datos conteniendo valores extremadamente altos de  $d$  estos índices podrían ser inaplicables para los estudios de similitud o asociación. Sin embargo, el mismo autor manifiesta que resulta sorprendente este resultado ya que esos mismos índices han sido

exitosamente utilizados en taxonomía numérica, la razón puede ser que en esta disciplina el valor de  $d$  suele ser generalmente bajo. Warrens (2008) también señala al coeficiente de Sokal & Michener como uno de los más recomendables.

Tabla 9: Monotonicidad de los índices de similitud para datos binarios. Correlaciones de Spearman y Kendall entre los índices y  $\chi^2$ .

Correlación de Spearman	Índices de similitud	Coeficientes de correlación	
		Spearman	Kendall
	Tschuproff	1,0000	1,0000
	Yule_1	1,0000	1,0000
	Cole_1	1,0000	1,0000
	Pearson_2	1,0000	1,0000
	Clifford & Stephenson	1,0000	1,0000
	Doolittle	1,0000	1,0000
	Pearson_1	1,0000	1,0000
	Maxwell & Pilliner	0,9993	0,9932
	Kendall	0,9943	0,9597
	Peirce_2	0,9900	0,9441
	Fleiss	0,9869	0,9491
	Cohen	0,9861	0,9363
<b>relativamente fuerte</b>	Michael	0,9801	0,9151
	Hurlbert_1	0,9781	0,9168
<b>0,90 ≤ r ≤ 1</b>	Dennis	0,9680	0,8760
	Forbes_2	0,9619	0,8678
	Yule_2	0,9505	0,8878
	Yule_3	0,9505	0,8878
	Loevinger	0,9503	0,8814
	Scott	0,9458	0,8644
	Hurlbert_2	0,9327	0,8317
	Goodman & Kruskal	0,9209	0,8083
	Baroni-Urbani & Buser_2	0,9173	0,7823
	Hamann	0,9091	0,8165
	Forbes_3	0,9054	0,7805
	Tarwid	0,9054	0,7805
	Mc Connaughey	0,9040	0,7570
	Cole_2	0,9010	0,8225
	Gilbert & Wells	0,8922	0,7655
	Anónimo_1	0,8371	0,6826
	Fager & Mc Gowan	0,8362	0,6698
	Kuder & Richardson	0,7115	0,6346
	Sokal & Sneath_2	0,7104	0,6338
	Peirce_3	0,7024	0,5970
	Mountford	0,6962	0,5929
	Sokal & Sneath_5	0,6895	0,5709
	Anónimo_5	0,6617	0,5185
	Anónimo_4	0,6550	0,5300
	Johnson_1	0,6520	0,5115
	Kulczynski_2	0,6520	0,5115
	Digby	0,6504	0,5251
	Hawkins & Dotson	0,6462	0,5259
	Harris & Lahey	0,6444	0,5142
	Rogot & Goldberg	0,6444	0,5235
	Clement	0,6436	0,4976
	Baroni-Urbani & Buser_1	0,6352	0,4949
<b>moderadamente fuerte</b>	Rogers & Tanimoto	0,6340	0,5188
	Goodall	0,6340	0,5188
<b>0,50 ≤ r &lt; 0,90</b>	Sokal & Michener	0,6340	0,5188
	Sokal & Sneath_4	0,6340	0,5188
	Cole_3	0,6214	0,4699
	Sokal & Sneath_3	0,6191	0,5044
	Sorgenfrei	0,6179	0,4719
	Ochiai	0,6179	0,4719
	Cole_4	0,6089	0,4569
	Dice_3	0,6029	0,4459
	Forbes_1	0,6029	0,4459
	Pearson & Heron	0,6014	0,4433
	Simpson	0,5867	0,4449
	Faith	0,5829	0,4450
	Sokal & Sneath_1	0,5797	0,4387
	Jaccard	0,5797	0,4387
	Lamont & Grant	0,5797	0,4387
	Sorensen	0,5797	0,4387
	Legendre & Legendre	0,5797	0,4387
	Fossum	0,5764	0,4322
	Kulczynski_1	0,5628	0,4229
	Anónimo_2	0,5594	0,4177
	Anónimo_3	0,5594	0,4177
	Stiles	0,5495	0,4431
<b>relativamente débil</b>	Braun-Blanquet	0,4895	0,3558
	Dice_2	0,4535	0,3205
<b>r &lt; 0,50</b>	Dice_1	0,4535	0,3205
	Russell & Rao	0,3631	0,2722
	Johnson_2	0,2244	0,1557
<b>no correlacionados</b>	Peirce_1	-0,0343	-0,0231
	Eyraud	-0,0106	0,0034

### **Propiedad 8: Linealidad**

Se estudia esta propiedad utilizando los escenarios planteados por Hubálek (1982) y Gower & Legendre (1986). Siguiendo el esquema de datos que utiliza Hubálek (1982), más de la mitad de los índices cumplen con la propiedad de linealidad (57%), (Tabla 10). Un 17% son cóncavos y un 25% son convexos. Por su parte, el índice de Pearson & Heron resulta cóncavo en dos ocasiones y convexo en otras dos.

Estos resultados coinciden con lo obtenido por Hubálek (1982), salvo para el índice de Cole\_1, que fue considerado cóncavo en su trabajo y convexo según los resultados obtenidos en esta Tesis, recordemos que existe un error en la fórmula de Hubálek. También el índice de Pearson & Heron es considerado convexo para ese autor mientras que según el estudio planteado resulta cóncavo y convexo según los escenarios. También en términos generales los resultados obtenidos coinciden con Giansanti (1990), sin embargo se discrepa con respecto al índice de Sorensen, que resulta convexo para ella mientras que es lineal para Hubálek (1982), Warrens (2008) y en este trabajo. Con respecto al índice de Mountford se coincide con Hubálek (1982) en su comportamiento como convexo, mientras que Giansanti (1990) lo considera cóncavo.

En la Tabla 10 se registran los índices agrupados según si cumplen o no la propiedad de linealidad, también se presentan los valores de Resolución, No-linealidad y CV descriptos en el Capítulo 2 (página 39). Gower & Legendre (1986) afirman que una alta resolución y una baja no-linealidad son adecuados, ambas condiciones derivan en un valor bajo de CV. La Tabla 10 está ordenada según el CV de menor a mayor, dentro de cada categoría: lineal, cóncavo y convexo. Entre los índices lineales se destaca que los índices de Dice\_3, Forbes\_1, Forbes\_3 y Peirce\_1 tienen un alto valor de CV. Según Gower y Legendre (1986) serían condiciones suficientes para descartarlos. En la condición opuesta, se destacan dentro de los índices cóncavos a Sokal y Sneath\_4 y dentro de los convexos a Jaccard y Sokal y Sneath\_5 como muy recomendables por los valores bajos de CV a pesar de no ser lineales.

Este estudio coincide con lo obtenido por Gower & Legendre (1986) quienes estudian la propiedad de linealidad en quince índices de similitud y encuentran que trece de ellos la cumplen: Jaccard, Kulczynski\_2, Hamann, Ochiai, Sokal &

Michener, Roger & Tanimoto, Russell & Rao, Sokal & Sneath\_1, Sokal & Sneath\_2, Sokal & Sneath\_4, Sokal & Sneath\_5, Sorensen y Yule\_1. Los índices de Kulczynski\_1 y Yule\_2 resultan no lineales para estos autores por sus altos valores de CV.

La Tabla 10 muestra que los índices de Clement, Fossum, Harris & Lahey y Pearson\_1 presentan una alta resolución, lo cual los hace apropiados para análisis cluster. También se desprende de la tabla que un amplio conjunto de índices con valores bajos de CV cumplen la propiedad de linealidad siendo por lo tanto aconsejables para Escalamiento métrico.

Tabla 10: Linealidad de los índices de similitud para datos binarios.

Linealidad	Índice de similitud	resolución	no-linealidad	CV
	Dice_1	0,098	0,018	18,23
	Peirce_3	0,114	0,023	20,37
	Maxwell & Pilliner	0,110	0,028	25,22
	Cohen	0,110	0,029	26,09
	Yule_1	0,109	0,029	26,40
	Tschuproff	0,104	0,028	26,78
	Rogot & Goldberg	0,061	0,017	27,14
	Scott	0,123	0,033	27,14
	Sokal & Sneath_2	0,053	0,015	28,57
	Faith	0,055	0,016	28,89
	Hamann	0,111	0,032	28,89
	Russell & Rao	0,055	0,016	28,89
	Sokal & Michener	0,055	0,016	28,89
	Fager & Mc Gowan	0,083	0,026	31,84
	Ochiai	0,081	0,026	31,95
	Anónimo_1	0,171	0,060	34,96
	Lamont & Grant	0,043	0,015	34,96
	Sorensen	0,086	0,030	34,96
	Johnson_1	0,146	0,053	36,18
	Kulczynski_2	0,073	0,026	36,18
<b>índices lineales</b>	Mc Connaughey	0,146	0,053	36,18
	Clement	8,844	3,236	36,59
	Cole_3	0,125	0,047	37,28
	Braun-Blanquet	0,081	0,031	38,58
	Goodman & Kruskal	0,165	0,064	38,91
	Peirce_2	0,099	0,040	40,87
	Forbes_2	0,101	0,044	43,40
	Fleiss	0,099	0,045	46,23
	Dennis	0,591	0,306	51,85
	Anónimo_2	0,037	0,020	52,66
	Anónimo_3	0,075	0,039	52,66
	Hurlbert_2	0,103	0,061	58,65
	Hurlbert_1	0,103	0,061	58,86
	Eyraud	0,000	0,000	65,21
	Anónimo_5	0,172	0,113	65,88
	Simpson	0,065	0,050	77,29
	Loevinger	0,096	0,075	78,19
	Cole_2	0,091	0,073	80,62
	Cole_4	0,072	0,062	86,65
	Dice_2	0,048	0,045	93,25
	Dice_3	0,001	0,001	104,80
	Forbes_1	0,089	0,094	104,80
	Forbes_3	0,089	0,094	104,80
	Peirce_1	0,020	0,021	109,00
	Sokal & Sneath_4	0,043	0,012	28,48
	Baroni-Urbani & Buser_1	0,075	0,023	31,24
	Baroni-Urbani & Buser_2	0,150	0,047	31,24
	Pearson_2	0,084	0,030	35,07
	Anónimo_4	0,074	0,029	38,58
	Goodall	0,040	0,016	39,10
<b>índices cóncavos</b>	Legendre & Legendre	0,083	0,040	48,20
	Kendall	0,076	0,040	52,60
	Kuder & Richardson	0,177	0,135	76,22
	Yule_3	0,091	0,071	77,18
	Stiles	1,118	1,027	91,86
	Tarwid	0,053	0,072	134,96
	Gilbert & Wells	0,115	0,164	143,01
	Jaccard	0,083	0,024	28,89
	Sokal & Sneath_5	0,076	0,023	29,71
	Sorgenfrei	0,087	0,028	31,97
	Fossum	12,919	4,210	32,58
	Harris & Lahey	10,228	3,488	34,10
	Sokal & Sneath_1	0,086	0,030	34,96
	Cole_1	0,125	0,046	37,14
	Hawkins & Dotson	0,064	0,025	39,84
<b>índices convexos</b>	Rogers & Tanimoto	0,062	0,028	44,87
	Michael	0,147	0,086	58,46
	Clifford & Stephenson	0,076	0,055	72,37
	Doolittle	0,076	0,055	72,37
	Pearson_1	11,464	8,296	72,37
	Digby	0,116	0,095	81,30
	Yule_2	0,133	0,117	88,32
	Johnson_2	0,741	1,147	154,80
	Mountford	0,016	0,027	167,35
	Kulczynski_1	0,773	1,377	178,27
	Sokal & Sneath_3	1,159	2,066	178,27
<b>cónc.y conv.</b>	Pearson & Heron	0,532	0,603	113,41

Nota: los sombreados corresponden a valores altos de resolución y a valores extremos de CV

### **Propiedades 9 y 10: Insensibilidad al valor de $a=0$ y Asociación positiva o negativa respectivamente**

La sensibilidad de un índice se identifica observando su variación frente a pequeños cambios en los valores de  $a$ ,  $b$ ,  $c$  y  $d$  manteniendo los totales marginales de la tabla de contingencias de  $2 \times 2$ .

Para identificar el cumplimiento de esta propiedad se grafican para cada índice los valores obtenidos en función de  $a$ , que varía de 0 a 40 en el caso considerado con  $n=100$  (Tabla 4 del Capítulo 2, página 51). Con cada cambio en  $a$  se modifican  $b$ ,  $c$  y  $d$  de modo de mantener constantes los totales marginales de la tabla de contingencia de  $2 \times 2$ .

Las representaciones logradas permiten también verificar visualmente la propiedad 8 de linealidad ya que cuando los índices son lineales la representación de los cambios sucesivos es aproximadamente una recta, con mayor o menor pendiente. Cuando los índices no cumplen la propiedad de linealidad y son cóncavos o convexos se obtienen curvas con concavidad hacia abajo o hacia arriba respectivamente, con curvaturas más o menos pronunciadas.

Estos mismos gráficos también se utilizan para verificar el cumplimiento de la propiedad 10, que se refiere a la posibilidad de identificar asociaciones positivas o negativas. Este aspecto está relacionado con el valor del índice frente a cambios en  $a$  (presencia conjunta). Se compara el valor de  $a$  con el de su esperanza ( $a'$ ) bajo el supuesto de independencia:

$$E(a) = a' = \frac{(a+b)(a+c)}{n}$$

Para los datos de la Tabla 4, el valor de  $a'=16$ , cualquiera sea el valor de  $a=0, 1, 2, 3, \dots, 40$ .

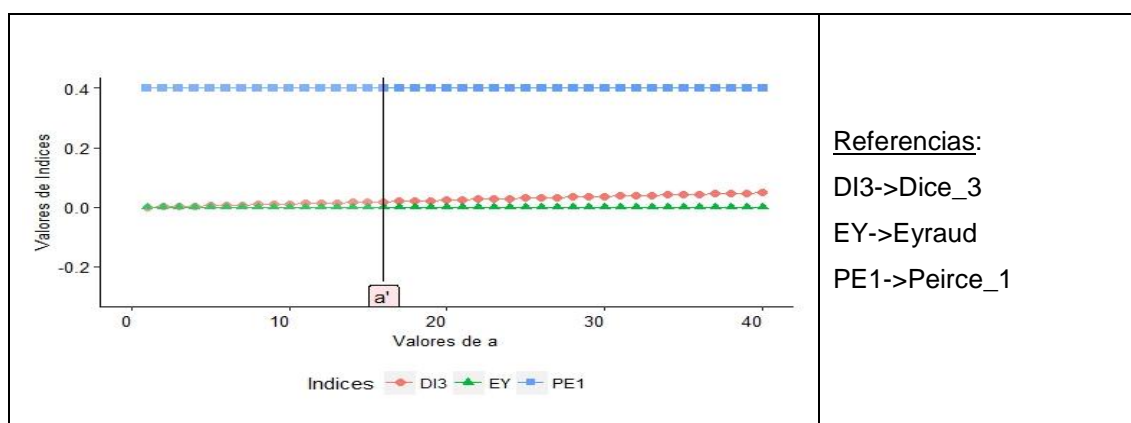
En esta sección las representaciones gráficas se exponen agrupando aquellos índices que tienen un comportamiento similar: constantes, lineales, cóncavos, convexos o mixtos.

Se encuentran 3 índices que son lineales y que se mantienen constantes a medida que el valor de  $a$  aumenta. Esos índices son Dice\_3, ligeramente

creciente y Eyraud y Peirce\_1 constantes. Esto implica el cumplimiento de la Propiedad 9.

La Propiedad 10 la cumple solamente Dice\_3, cuya curva es levemente creciente. La Figura 1 muestra que cuando se considera un valor  $a_1 < a' = 16$  y otro valor  $a_2 > a'$  entonces la similitud  $A_i(a_1)$  es siempre menor que  $A_i(a_2)$ .

Figura 1: Propiedad de insensibilidad al valor de  $a=0$  y asociación positiva o negativa. Tendencia constante. Rango de variación  $[-1,1]$ .



Referencias:  
 DI3->Dice\_3  
 EY->Eyraud  
 PE1->Peirce\_1

Catorce índices que varían entre 0 y 1 y son lineales, se grafican por separado para identificar con mayor claridad las rectas (Fig. 2a y 2b). Las pendientes de las rectas resultan un indicador de la sensibilidad respecto de  $a=0$ , a menor pendiente es más evidente el cumplimiento de la propiedad 9. En la Fig. 2a, el índice menos sensible al valor de  $a=0$  es el de Russell & Rao, luego le sigue Lamont y por último Faith. En la Fig.2b el índice menos sensible es Anónimo\_2, luego están los índices de Sokal & Sneath\_2 y Rogot & Goldberg que coinciden y Sokal & Michener que tiene también una línea casi idéntica a la de los anteriores, por último hay siete índices que coinciden: Sorensen, Dice\_1, Simpson, Dice\_2, Ochiai, Kulczynski\_2 y Braun-Blanquet.

Todos los índices, al ser lineales crecientes cumplen la Propiedad 10.

En las Figuras 2c y 2d se muestran dieciséis índices que varían entre -1 y 1, que son lineales y se grafican por grupos para diferenciar las rectas que muestran su variación frente a cambios consecutivos de  $a$ . El índice menos sensible es Fager & Mc Gowan, le sigue Anónimo\_3 y luego Haman (Fig. 2c). En la Fig.2d aparecen

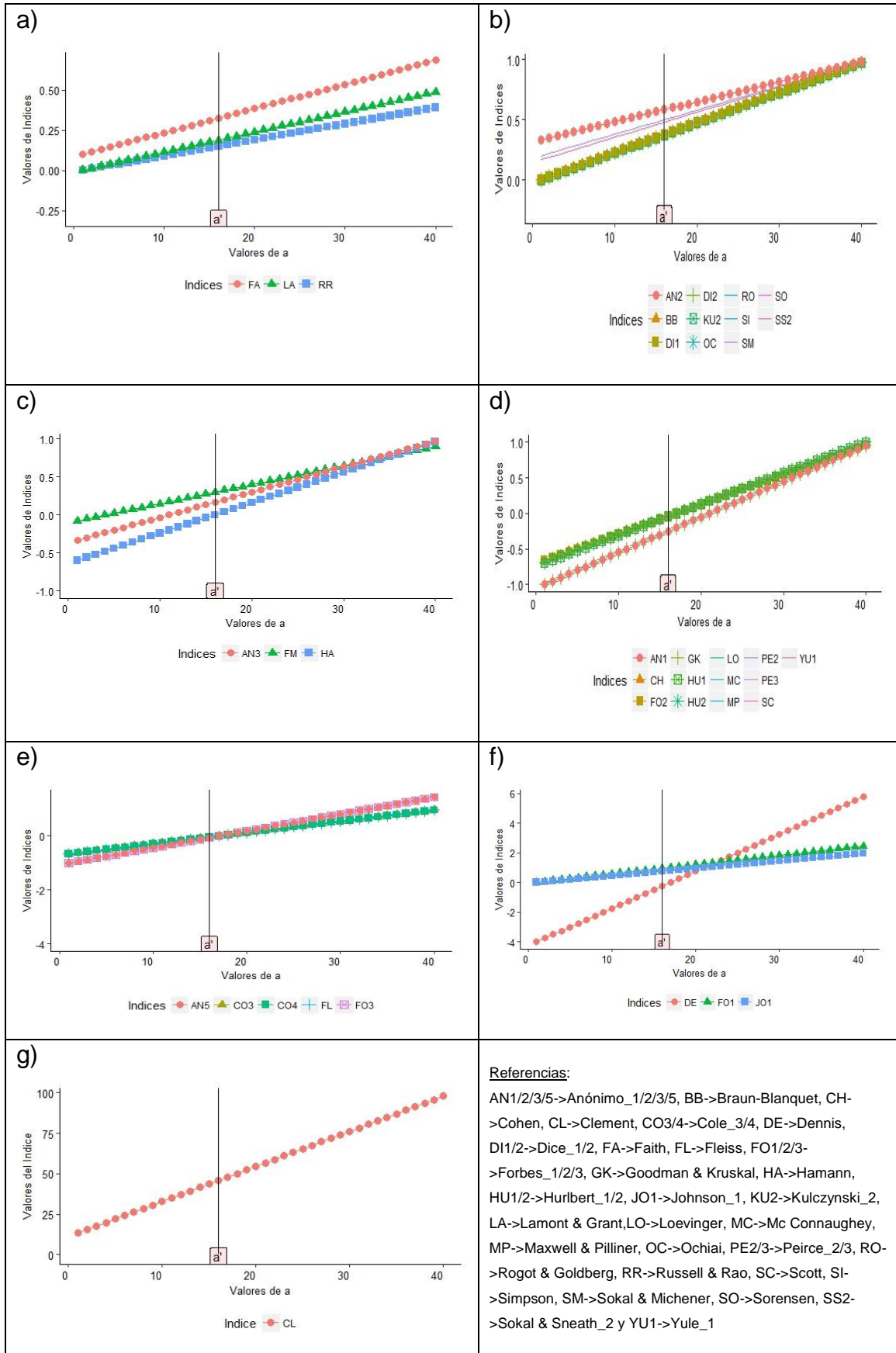
dos grupos de índices con comportamientos ligeramente distintos. Uno de ellos, que muestra una recta con menos pendiente en su variación desde  $a=0$  hasta  $a=40$  está conformado por los índices de Yule\_1, Scott, Peirce\_2, Maxwell & Pilliner, Loevinger, Hurlbert\_2, Hurlbert\_1, Forbes\_2, Peirce\_3 y Cohen. El otro grupo lo forman Anónimo\_1, Mc Connaughey y Goodman & Kruskal. Todos los índices, al ser lineales crecientes cumplen la Propiedad 10.

En las Figuras 2e y 2f se tienen ocho índices que varían entre  $-\infty$  y  $+\infty$ , que son lineales y se han graficado por separado para identificar más fácilmente su comportamiento. En la Fig.2e los índices de Fleiss, Cole\_3 y Cole\_4 coinciden entre si y tienen rectas muy cercanas a Forbes\_3 y Anónimo\_5 que también coinciden. En la Fig.2f los índices de Forbes\_1 y Johnson\_1 también coinciden. Los anteriores muestran menor cambio al aumentar el valor  $a=0$ . Por su parte el índice de Dennis crece más rápido.

Todos los índices, al ser lineales crecientes cumplen la Propiedad 10.

En la Figura 2g se aprecia el comportamiento del índice de Clement, que varía entre 0 y  $+\infty$  y es lineal, que también tiene cambios graduales y cumple la propiedad 10.

Figura 2: Propiedad de insensibilidad al valor de  $a=0$  y asociación positiva o negativa. Tendencia creciente lineal.



**Referencias:**  
 AN1/2/3/5->Anónimo\_1/2/3/5, BB->Braun-Blanquet, CH->Cohen, CL->Clement, CO3/4->Cole\_3/4, DE->Dennis, DI1/2->Dice\_1/2, FA->Faith, FL->Fleiss, FO1/2/3->Forbes\_1/2/3, GK->Goodman & Kruskal, HA->Hamann, HU1/2->Hurlbert\_1/2, JO1->Johnson\_1, KU2->Kulczynski\_2, LA->Lamont & Grant, LO->Loevinger, MC->Mc Connaughey, MP->Maxwell & Pilliner, OC->Ochiai, PE2/3->Peirce\_2/3, RO->Rogot & Goldberg, RR->Russell & Rao, SC->Scott, SI->Simpson, SM->Sokal & Michener, SO->Sorensen, SS2->Sokal & Sneath\_2 y YU1->Yule\_1

Otro grupo de índices Johnson\_2 y Kulczynski\_1, que coinciden, y Sokal & Sneath\_3 (Fig. 3a) y Mountford (Fig. 3b) se mantienen más o menos constantes para valores pequeños de  $a$  y para valores grandes, mayores que 30 aproximadamente, en los casos considerados, la curva presenta un crecimiento abrupto. De este modo los índices son sensibles al valor de  $a=0$  y convexos. El índice de Mountford, fue graficado aparte por tener diferente rango de variación con respecto a los tres anteriores.

Todos los índices, al ser curvos crecientes cumplen la Propiedad 10.

Otro conjunto de seis índices que varían entre 0 y 1, son convexos y se consideran insensibles al valor de  $a=0$  son: Hawkins & Dotson y Roger & Tanimoto que coinciden, Sokal & Sneath\_5 y Jaccard que también coinciden, Sorgenfrei y Sokal & Sneath\_1. De ellos el menos sensible es el índice de Sokal & Sneath\_1. (Figura 3c).

Todos los índices, al ser curvos crecientes cumplen la Propiedad 10.

En la Figura 3d se encuentran los índices que varían entre -1 y 1 y que presentan también alguna curvatura, y que son convexos. Se encuentran bastante superpuestos los índices de Cole\_1, Digby, Michael y Yule\_2 y el cambio es gradual.

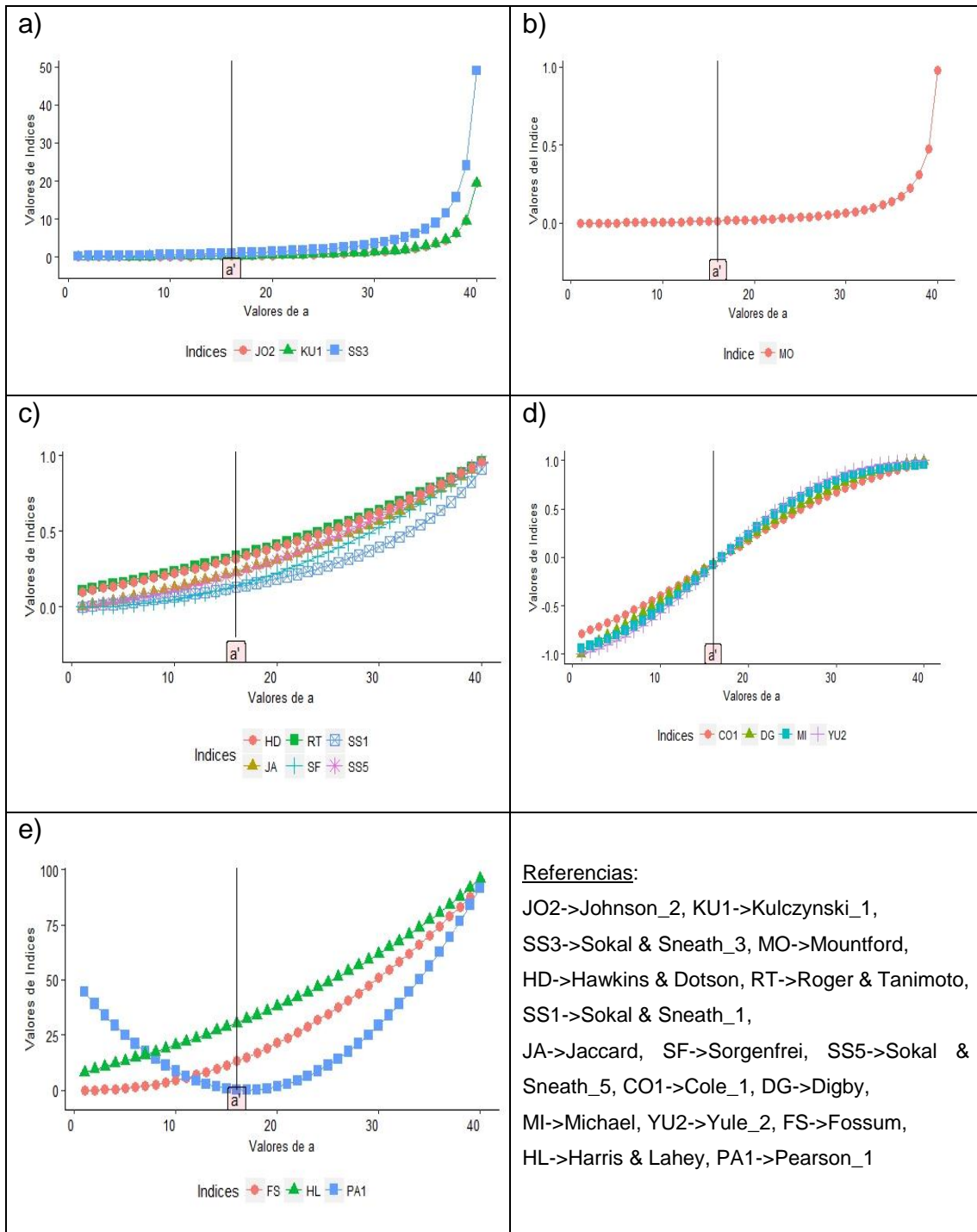
Las curvas de cambios sucesivos de estos índices muestran un leve cambio de inflexión en  $a'$ , sin embargo, de acuerdo a los cálculos de este trabajo resultan todos convexos. Cole\_1 es considerado cóncavo por Hubálek, sin embargo existe un error en la fórmula reportada por el autor.

Todos los índices, al ser curvos crecientes cumplen la Propiedad 10.

En la Figura 3e se observan tres índices que varían entre 0 y  $+\infty$  y presentan un comportamiento curvilíneo, Harris & Lahey, Fossum y Pearson\_1, todos convexos, este último tiene en el numerador el cuadrado de  $(ad-bc)$  sin considerar los casos negativos, de aquí esa forma. Los índices de Harris & Lahey y Fossum muestran cambios graduales mientras que Pearson\_1 no.

Solamente el índice de Pearson\_1 no cumple la Propiedad 10.

Figura 3: Propiedad de insensibilidad al valor de  $a=0$  y asociación positiva o negativa. Tendencia convexa creciente.



**Referencias:**

JO2->Johnson\_2, KU1->Kulczynski\_1,  
 SS3->Sokal & Sneath\_3, MO->Mountford,  
 HD->Hawkins & Dotson, RT->Roger & Tanimoto,  
 SS1->Sokal & Sneath\_1,  
 JA->Jaccard, SF->Sorgenfrei, SS5->Sokal &  
 Sneath\_5, CO1->Cole\_1, DG->Digby,  
 MI->Michael, YU2->Yule\_2, FS->Fossum,  
 HL->Harris & Lahey, PA1->Pearson\_1

Entre aquellos índices que varían entre 0 y 1, presentan una leve concavidad hacia abajo (cóncavos) y se los considera insensibles al valor de  $a=0$  se tienen: Anónimo\_4, Baroni-Urbani\_1 y Legendre & Legendre, muy juntos y casi superpuestos, Goodall y Sokal & Sneath\_4 (Fig. 4a). Sin embargo, entre ellos se observa una pequeña diferencia siendo Goodall el menos sensible.

Todos los índices, al ser curvos crecientes cumplen la Propiedad 10.

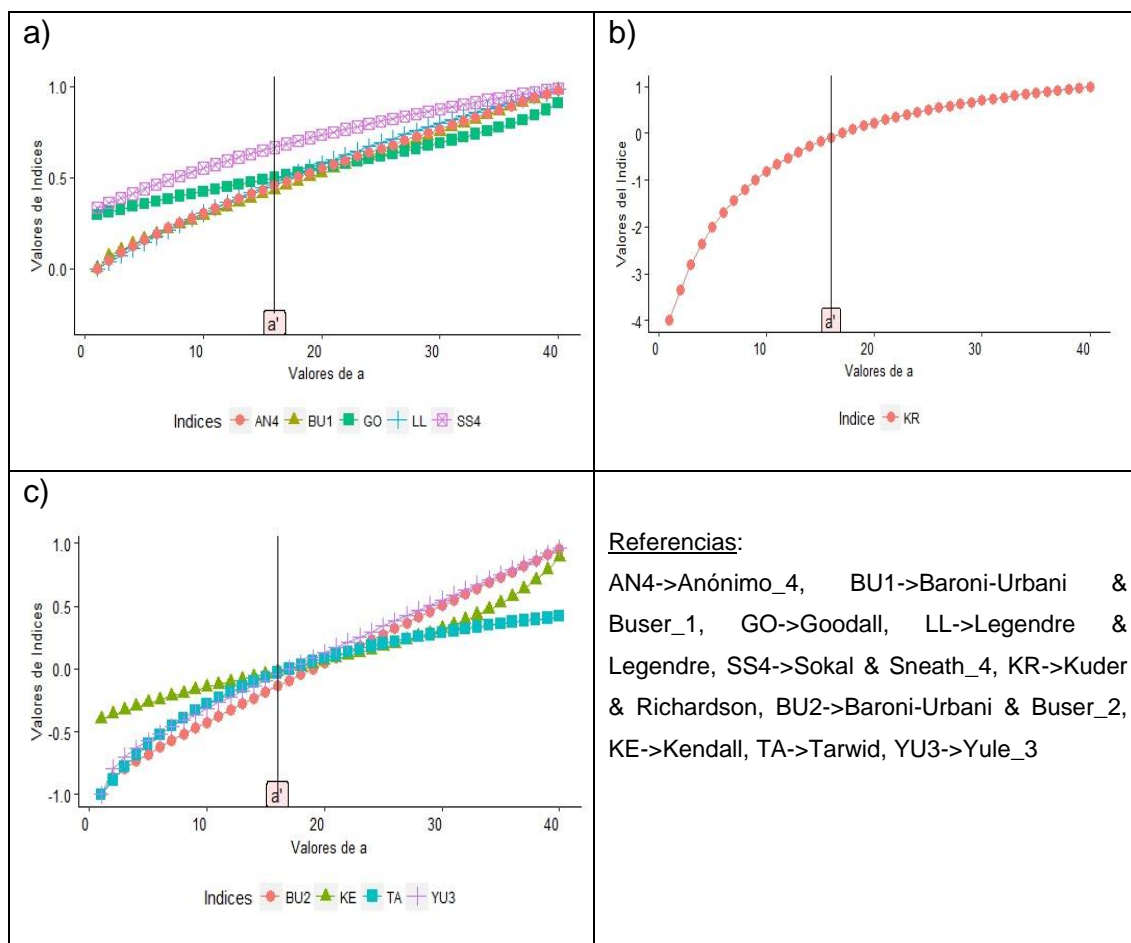
El índice de Kuder & Richardson (Figura 4b) presenta una curvatura muy pronunciada, es cóncavo y no cumple esta propiedad. Se grafica aparte por tener otro rango de variación.

Si cumple con la Propiedad 10 al ser curvo creciente.

En la Figura 4c se estudian los índices que varían entre -1 y 1 y presentan alguna curvatura, más precisamente son cóncavos. El índice con concavidad menos pronunciada es el de Tarwid. Los restantes Kendall, Baroni-Urbani\_2 y Yule\_3 si bien no presentan cambios uniformes frente a  $a$  son aproximadamente equivalentes en cuanto a la magnitud de cambio en todo el rango de variación de  $a$ .

Todos los índices cumplen con la Propiedad 10.

Figura 4: Propiedad de insensibilidad al valor de  $a=0$  y asociación positiva o negativa. Tendencia cóncava creciente.



En la Figura 5a se grafican los cambios para cinco índices que varían entre 0 y 1 evidenciando el no cumplimiento de la propiedad de insensibilidad al valor de  $a=0$ . Estos presentan un comportamiento parecido al variar el valor de  $a=0$ , primero descendente hasta la mitad del rango considerado y luego ascendente en la otra mitad. Este comportamiento se debe a que los índices por definición son siempre positivos, aun cuando la expresión de sus numeradores  $ad-bc$  sea negativa. Dos de estos índices Pearson\_2 y Tschuproff consideran la raíz positiva de una expresión donde  $ad-bc$  interviene al cuadrado. Por su parte el índice de Clifford-Stephenson corresponde al  $\chi^2$  estandarizado por el valor de  $n$  y es siempre positivo. Lo mismo ocurre con el índice de Cole\_2, que es el valor de  $\Phi$  estandarizado por el valor de  $\Phi_{\text{máx}}$  con el mismo signo de  $\Phi$ , por lo tanto es siempre positivo. Por su parte el índice de Doolittle es el valor de  $\Phi$  al cuadrado. En la Figura 5b se encuentra el índice de Stiles que tiene un

comportamiento parecido a los anteriores, ya que en su fórmula interviene el valor absoluto de  $ad-bc$ . Se grafica aparte porque varía entre  $-\infty$  y  $+\infty$ .

Si el signo de  $ad-bc$  fuese tenido en cuenta, estos segmentos decrecientes serían negativos y crecientes, es decir, estarían rebatidos en el gráfico respecto al eje horizontal, como ocurre con los índices de Hurlbert\_1 y Hurlbert\_2, los cuales son poco sensibles al valor de  $a=0$  (Fig. 2d).

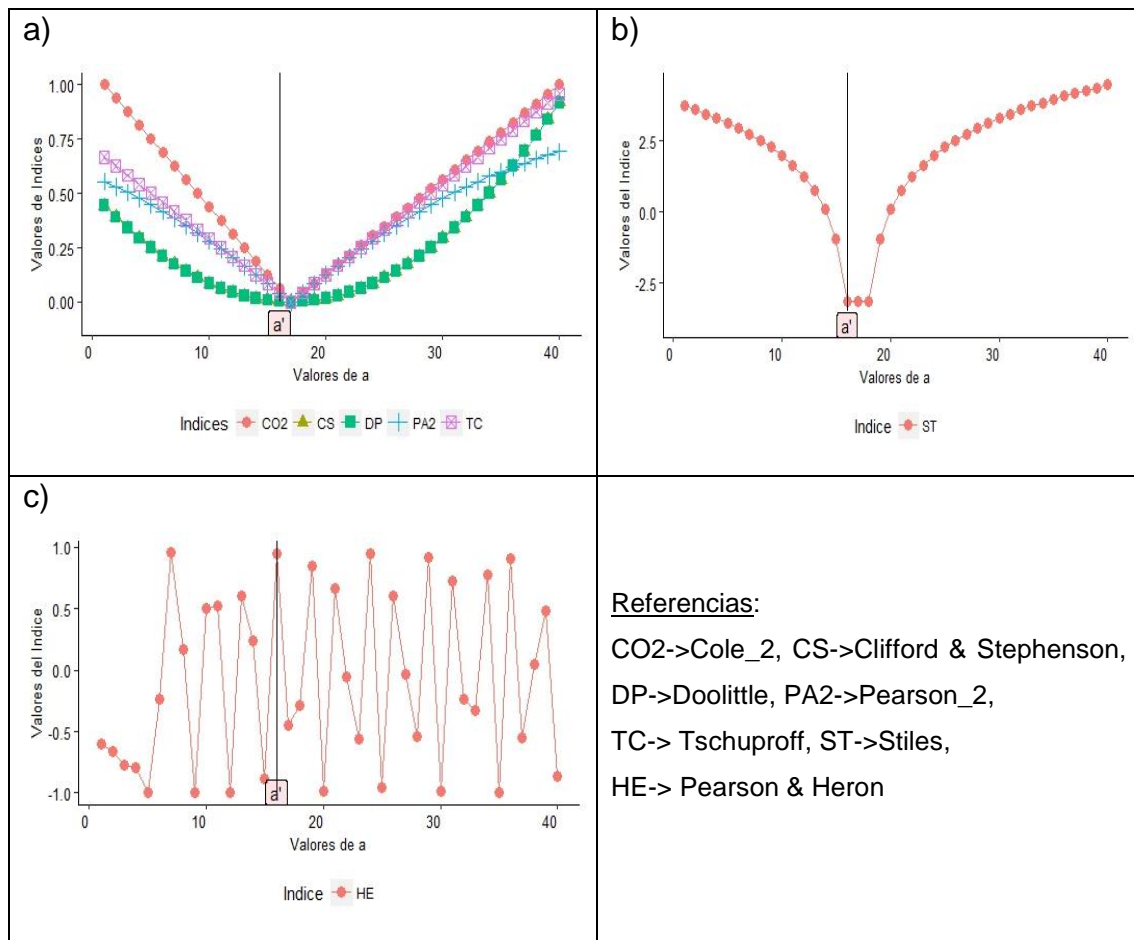
Ninguno de los índices cumple con la Propiedad 10.

Cabe aclarar que si bien estos seis índices tienen un comportamiento parecido, con respecto a la linealidad presentan comportamientos disimiles. Los índices de Cole\_2 y Tschuproff son lineales, mostrando una forma de V. Los índices de Clifford & Stephenson y Doolittle, que coinciden son convexos, mostrando concavidad hacia arriba, mientras que los índices de Pearson\_2 y Stiles son cóncavos.

En la Figura 5c se observa el comportamiento oscilante del índice de Pearson & Heron, que resulta lógico ya que en su fórmula hay un coseno, mostrando cambios bruscos frente a cambios unitarios de  $a$ .

El índice de Pearson & Heron por tener un comportamiento creciente y decreciente no cumple con la Propiedad 10.

Figura 5: Propiedad de insensibilidad al valor de  $a=0$  y asociación positiva o negativa. Tendencia cóncava/convexa/lineal.



En resumen, para la Propiedad 9, luego de haber estudiado los cambios en el valor de los índices  $A_i(\mathbf{j}, \mathbf{k})^{(a,b,c,d)}$  cuando el valor de  $a=0$ , es decir  $A_i(\mathbf{j}, \mathbf{k})^{(a=0,b,c,d)}$  comparado con la secuencia de valores  $A_i(\mathbf{j}, \mathbf{k})^{(a=1,b-1,c-1,d+1)}$ ,  $A_i(\mathbf{j}, \mathbf{k})^{(a=2,b-2,c-2,d+2)}$ ,  $A_i(\mathbf{j}, \mathbf{k})^{(a=3,b-3,c-3,d+3)}$ , ..., destacamos que la mayoría de los índices (82%) tiene un comportamiento gradual, ya sea lineal o curvilíneo. Los índices que tienen un cambio considerable y que por lo tanto no cumplen la propiedad son Johnson\_2, Kulczynski\_1, Sokal & Sneath\_3, Mountford, Kuder & Richardson, Cole\_2, Clifford & Stephenson, Doolittle, Pearson\_2, Tschuproff, Stiles, Pearson\_1 y Person & Heron (17%). Por su parte para el índice de Gilbert & Wells no se puede evaluar esta propiedad ya que el mismo no está definido para  $a=0$ .

El índice de Pearson\_1 fue indicado por Hubálek (1982) como sensible, también los índices de Baroni-Urbani\_1, Baroni-Urbani\_2 y Yule\_3, pero en este trabajo

resultan igualmente insensibles que otros índices que se estudian en esta tesis y que no fueron considerados por Hubálek.

En resumen, para la Propiedad 10, cuando se evalúan los cambios de  $A_i$  a medida que los valores de las presencias conjuntas varían desde  $a=0$  hasta  $a=40$ , manteniendo los mismos totales marginales, teniendo como referencia el valor de  $a'=16$ , se observa que, para los índices con curva creciente, tanto lineal como ligeramente curva se cumple esta propiedad con alta frecuencia (87%). No se cumple para los índices de Clifford & Stephenson, Cole\_2, Doolittle, Eyraud, Pearson\_2, Tschuproff, Stiles, Pearson\_1, Pearson & Heron y Peirce\_1 (13%).

Hubálek (1982) reporta que los índices de Doolittle, Pearson\_2 y Pearson\_1 también son incompatibles con la propiedad 10.

### **Propiedad 11: Distribuciones en el muestreo**

Se aproximan las distribuciones de todos los índices de similitud ( $A_i$ ) y de los tres índices de disimilitud derivados de los anteriores,  $D1_i$ ,  $D2_i$  y  $D3_i$ , a partir de un procedimiento de simulación descrito en la sección 2. 5. (página 49).

Se construyen histogramas (Figura 6), se calculan medidas descriptivas de asimetría y curtosis (Tabla 11) y se prueba la normalidad utilizando los tests de Shapiro-Wilks y Kolmogorov-Smirnov. Se verifica si las distribuciones empíricas presentan asimetría y curtosis similar a la distribución Normal mediante dos procedimientos. Uno de ellos consiste en calcular la asimetría y curtosis estandarizada verificando si las mismas caen dentro del intervalo  $[-2, 2]$ . El otro criterio consiste en utilizar el gráfico de Cullen y Frey (1999) que superpone la distribución empírica con una variedad de modelos con asimetría y curtosis iguales a ella, verificando su cercanía a una Normal (Tabla 11).

No se encontraron índices cuyas distribuciones cumplan con el supuesto de normalidad ( $p < 0,001$ ), si se encontraron algunos que tienen coeficientes de asimetría y curtosis acorde a una distribución Normal. Para algunas distribuciones empíricas la asimetría/curtosis se superpone a la asimetría y curtosis de una distribución Normal, en el sentido del gráfico de Cullen y Frey. (Tabla 11).

En la mayoría de los casos estudiados el ajuste normal no se verifica, a lo sumo coinciden solo algunas características muestrales como asimetría y curtosis. Algunas de las distribuciones se asemejan a otros modelos de probabilidad distintos de la normal, como distribuciones uniformes, exponencial, lognormal, beta, etc.

La Tabla 11 presenta estos resultados agrupando a los índices según su rango de variación. Dentro de cada grupo se ordenan alfabéticamente por autores, El mismo ordenamiento se respeta en la Figura 6 que presenta las distribuciones empíricas obtenidas.

Tabla 11: Coeficientes de asimetría y curtosis de las distribuciones en el muestreo de los índices de similitud (A) y disimilitud (D1, D2 y D3).

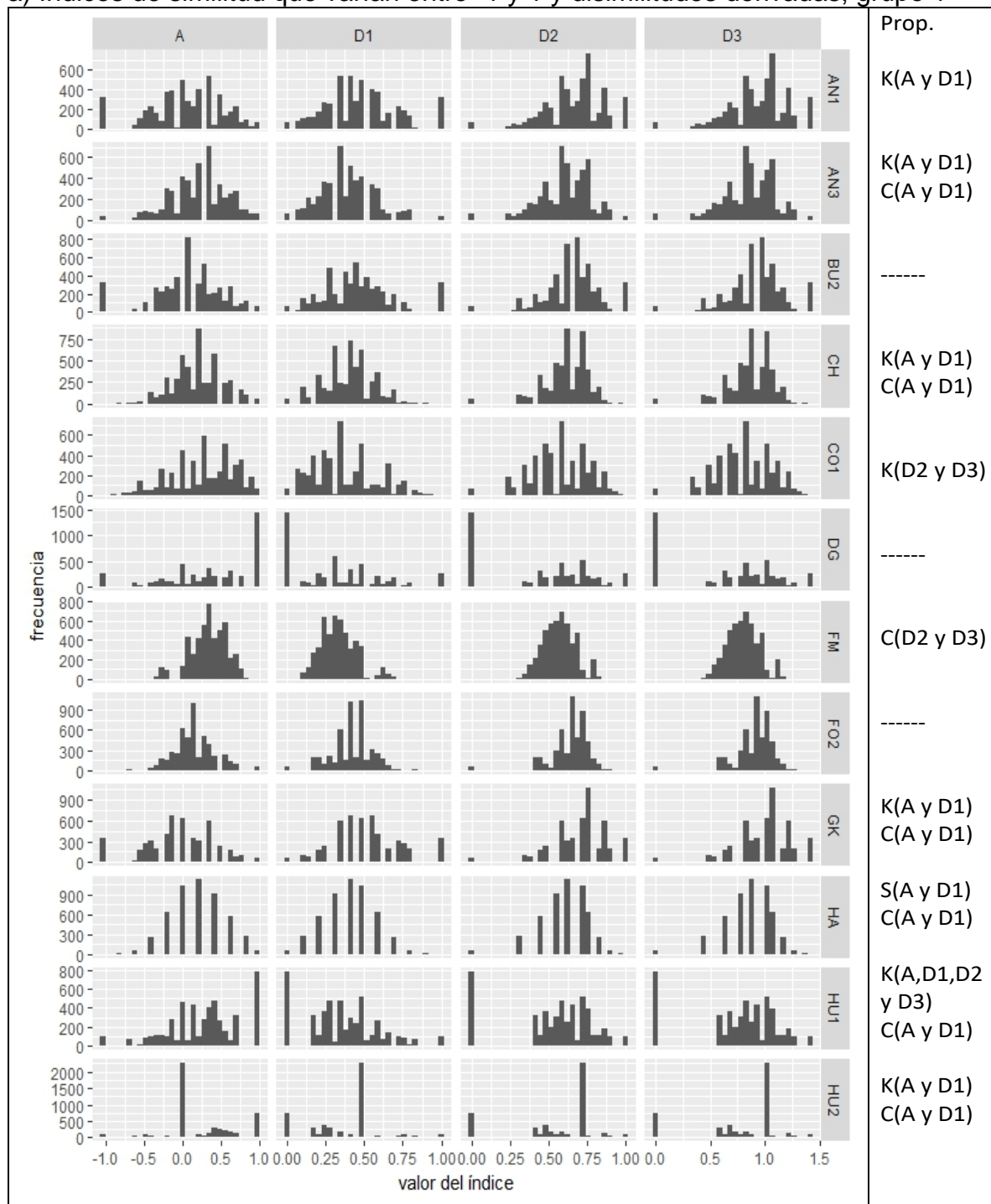
Rango de variación	Índices de similitud	Abreviatura	Coeficiente de asimetría				Coeficiente de curtosis				Aprox.Normal (*)			
			A	D1	D2	D3	A	D1	D2	D3	A	D1	D2	D3
	Anónimo_1	AN1	-12,08	12,08	-17,22	-17,22	-1,27	-1,27	16,22	16,22	no	no	no	no
	Anónimo_3	AN3	-9,70	9,70	-20,65	-20,65	1,09	1,09	20,95	20,95	si	si	no	no
	Baroni-Urbani & Buser_2	BU2	-17,64	17,64	-15,78	-15,78	6,93	6,93	22,44	22,44	no	no	no	no
	Cohen	CH	2,46	-2,46	-36,31	-36,31	-0,89	-0,89	48,45	48,45	si	si	no	no
	Cole_1	CO1	-12,11	12,11	-10,56	-10,56	-8,12	-8,12	0,39	0,39	no	no	no	no
	Digby	DG	-16,69	16,69	-9,03	-9,03	-5,64	-5,64	-17,85	-17,85	no	no	no	no
	Fager & Mc Gowan	FM	-16,38	16,38	2,69	2,69	3,75	3,75	-2,11	-2,11	no	no	si	si
	Forbes_2	FO2	12,97	-12,97	-60,27	-60,27	11,24	11,24	123,88	123,88	no	no	no	no
grupo 1	Goodman & Kruskal	GK	-6,68	6,68	-26,60	-26,60	-0,72	-0,72	34,91	34,91	si	si	no	no
[-1,1]	Hamann	HA	0,82	-0,82	-30,90	-30,90	-4,38	-4,38	33,50	33,50	si	si	no	no
	Hurlbert_1	HU1	-8,58	8,58	-26,40	-26,40	-1,30	-1,30	0,41	0,41	si	si	no	no
	Hurlbert_2	HU2	2,41	-2,41	-32,45	-32,45	-0,93	-0,93	3,40	3,40	si	si	no	no
	Kendall	KE	27,02	-27,02	-92,06	-92,06	40,79	40,79	245,59	245,59	no	no	no	no
	Maxwell & Pilliner	MP	-3,06	3,06	-27,33	-27,33	-4,33	-4,33	30,83	30,83	si	si	no	no
	Mc Connaughey	MC	-18,65	18,65	-11,55	-11,55	4,90	4,90	13,80	13,80	no	no	no	no
	Michael	MI	-14,59	14,59	-5,56	-5,56	-8,34	-8,34	-8,50	-8,50	no	no	no	no
	Pearson & Heron	HE	-14,75	14,75	-2,11	-2,11	-18,10	-18,10	-20,63	-20,63	no	no	no	no
	Peirce_2	PE2	-4,98	4,98	-25,44	-25,44	-2,57	-2,57	26,53	26,53	si	si	no	no
	Peirce_3	PE3	-3,17	3,17	-26,21	-26,21	-3,77	-3,77	25,34	25,34	si	si	no	no
	Scott	SC	2,99	-2,99	-33,95	-33,95	-4,52	-4,52	39,75	39,75	si	si	no	no
	Tarwid	TA	-62,29	62,29	42,45	42,45	88,39	88,39	57,62	57,62	no	no	no	no
	Yule_1	YU1	-3,54	3,54	-26,53	-26,53	-4,53	-4,53	29,28	29,28	si	si	no	no
	Yule_2	YU2	-19,68	19,68	-4,72	-4,72	-7,38	-7,38	-18,66	-18,66	no	no	no	no
	Yule_3	YU3	-10,25	10,25	-15,65	-15,65	-3,64	-3,64	-16,63	-16,63	no	no	no	no
	Anonimo_2	AN2	-9,70	9,70	-20,65	-20,65	1,09	1,09	20,95	20,95	si	si	no	no
	Anonimo_4	AN4	-17,24	17,24	-17,08	-17,08	6,97	6,97	26,02	26,02	no	no	no	no
	Baroni-Urbani & Buser_1	BU1	-17,64	17,64	-15,78	-15,78	6,93	6,93	22,44	22,44	no	no	no	no
	Braun-Blanquet	BB	-0,14	0,14	-32,02	-32,02	-3,99	-3,99	40,73	40,73	si	si	no	no
	Clifford & Stephenson	CS	54,01	-54,01	-102,70	-102,70	56,19	56,19	249,47	249,47	no	no	no	no
	Cole_2	CO2	5,84	-5,84	-14,26	-14,26	-21,58	-21,58	-22,57	-22,57	no	no	no	no
	Dice_1	DI1	-3,77	3,77	-26,12	-26,12	-9,12	-9,12	-2,10	-2,10	no	no	no	no
	Dice_2	DI2	-2,61	2,61	-28,23	-28,23	-9,07	-9,07	1,74	1,74	no	no	no	no
	Doolittle	DP	54,01	-54,01	-102,70	-102,70	56,19	56,19	249,47	249,47	no	no	no	no
	Faith	FA	14,31	-14,31	-23,52	-23,52	-0,39	-0,39	9,43	9,43	no	no	no	no
	Goodall	GO	14,99	-14,99	-70,54	-70,54	16,86	16,86	167,87	167,87	no	no	no	no
	Hawkins & Dotson	HD	21,60	-21,60	-61,62	-61,62	7,34	7,34	110,67	110,67	no	no	no	no
	Jaccard	JA	10,82	-10,82	-45,50	-45,50	-1,24	-1,24	65,86	65,86	no	no	no	no
	Kulczynski_2	KU2	-20,49	20,49	-8,33	-8,33	3,15	3,15	10,60	10,60	no	no	no	no
	Legendre & Legendre	LL	-26,56	26,56	-3,17	-3,17	10,33	10,33	7,08	7,08	no	no	si	si
	Ochiai	OC	-16,17	16,17	-13,53	-13,53	1,38	1,38	14,32	14,32	no	no	no	no
grupo 2	Pearson_2	PA2	4,93	-4,93	-12,09	-12,09	-12,12	-12,12	-8,23	-8,23	no	no	no	no
[0,1]	Peirce_1	PE1	6,99	-6,99	-36,83	-36,83	-9,11	-9,11	11,21	11,21	no	no	no	no
	Rogers & Tanimoto	RT	18,55	-18,55	-55,78	-55,78	3,25	3,25	92,04	92,04	no	no	no	no
	Rogot & Goldbert	RO	2,99	-2,99	-33,95	-33,95	-4,52	-4,52	39,75	39,75	si	si	no	no
	Russell & Rao	RR	14,00	-14,00	-22,93	-22,93	-0,86	-0,86	8,47	8,47	no	no	no	no
	Simpson	SI	-21,19	21,19	-8,70	-8,70	-0,20	-0,20	-16,53	-16,53	no	no	no	no
	Sokal & Michener	SM	0,82	-0,82	-30,90	-30,90	-4,38	-4,38	33,50	33,50	si	si	no	no
	Sokal & Sneath_1	SS1	35,60	-35,60	-85,46	-85,46	29,22	29,22	194,02	194,02	no	no	no	no
	Sokal & Sneath_2	SS2	-4,25	4,25	-25,53	-25,53	-4,67	-4,67	27,63	27,63	si	si	no	no
	Sokal & Sneath_4	SS4	-16,07	16,07	-13,91	-13,91	6,74	6,74	12,63	12,63	no	no	no	no
	Sokal & Sneath_5	SS5	15,19	-15,19	-51,27	-51,27	0,89	0,89	80,15	80,15	no	no	no	no
	Sorensen	SO	-12,08	12,08	-17,22	-17,22	-1,27	-1,27	16,22	16,22	no	no	no	no
	Sorgenfrei	SF	15,61	-15,61	-46,65	-46,65	-3,84	-3,84	61,06	61,06	no	no	no	no
	Tschuproff	TC	17,19	-17,19	-49,19	-49,19	-2,51	-2,51	68,62	68,62	no	no	no	no
grupo 3	Dice_3	DI3	29,58	-29,58	-39,24	-39,24	69,21	69,21	95,39	95,39	no	no	no	no
[0,2]	Johnson_1	JO1	-18,92	18,92	-10,66	-10,66	4,39	4,39	12,78	12,78	no	no	no	no
	Mountford	MO	61,52	-61,52	-90,04	-90,04	67,47	67,47	134,46	134,46	no	no	no	no
grupo 4	Clement	CL	-7,43	7,43	-17,71	-17,71	-3,41	-3,41	3,35	3,35	si	si	no	no
[0,n]	Harris & Lahey	HL	20,49	-20,49	-62,03	-62,03	7,91	7,91	114,73	114,73	no	no	no	no
	Pearson_1	PA1	54,01	-54,01	-102,70	-102,70	56,19	56,19	249,47	249,47	no	no	no	no
	Cole_3	CO3	-1,23	1,23	-34,03	-34,03	12,79	12,79	5,49	5,49	no	no	no	no
grupo 5	Cole_4	CO4	-14,70	14,70	-34,95	-34,95	64,10	64,10	12,60	12,60	no	no	no	no
[-∞,1]	Fleiss	FL	-12,91	12,91	-13,63	-13,63	2,43	2,43	9,08	9,08	no	no	no	no
	Kuder & Richardson	KR	-128,68	128,68	31,47	31,47	586,22	586,22	60,94	60,94	no	no	no	no
	Loevinger	LO	-19,73	19,73	-13,58	-13,58	24,45	24,45	-15,91	-15,91	no	no	no	no
grupo 6														
[-1,0]	Eyraud	EY	-67,42	67,42	-70,75	-70,75	129,50	129,50	144,61	144,61	no	no	no	no
grupo 7														
[0,0.5]	Lamont & Grant	LA	-12,08	12,08	-17,22	-17,22	-1,27	-1,27	16,22	16,22	no	no	no	no
	Forbes_1	FO1	29,58	-	-	-	69,21	-	-	-	no	-	-	-
grupo 8	Fossum	FS	23,61	-	-	-	0,49	-	-	-	no	-	-	-
[0,+∞]	Johnson_2	JO2	39,29	-	-	-	26,99	-	-	-	no	-	-	-
	Kulczynski_1	KU1	83,80	-	-	-	148,99	-	-	-	no	-	-	-
	Sokal & Sneath_3	SS3	68,15	-	-	-	80,80	-	-	-	no	-	-	-
grupo 9	Anonimo_5	AN5	19,16	-	-	-	34,53	-	-	-	no	-	-	-
[-1,+∞]	Forbes_3	FO3	29,58	-	-	-	69,21	-	-	-	no	-	-	-
	Dennis	DE	2,46	-	-	-	4,33	-	-	-	si	-	-	-
grupo 10	Gilbert & Wells	GI	1,44	-	-	-	15,33	-	-	-	no	-	-	-
[-∞,+∞]	Stiles	ST	-18,45	-	-	-	-3,11	-	-	-	no	-	-	-

(\*) según el criterio de Cullen y Frey

Nota: el sombreado corresponde a la observación de pareado de asimetría y/o curtosis de la distribución normal

Figura 6: Distribuciones en el muestreo de los índices de similitud (A) y disimilitud (D1, D2 y D3) agrupados por rango de variación de la similitud.

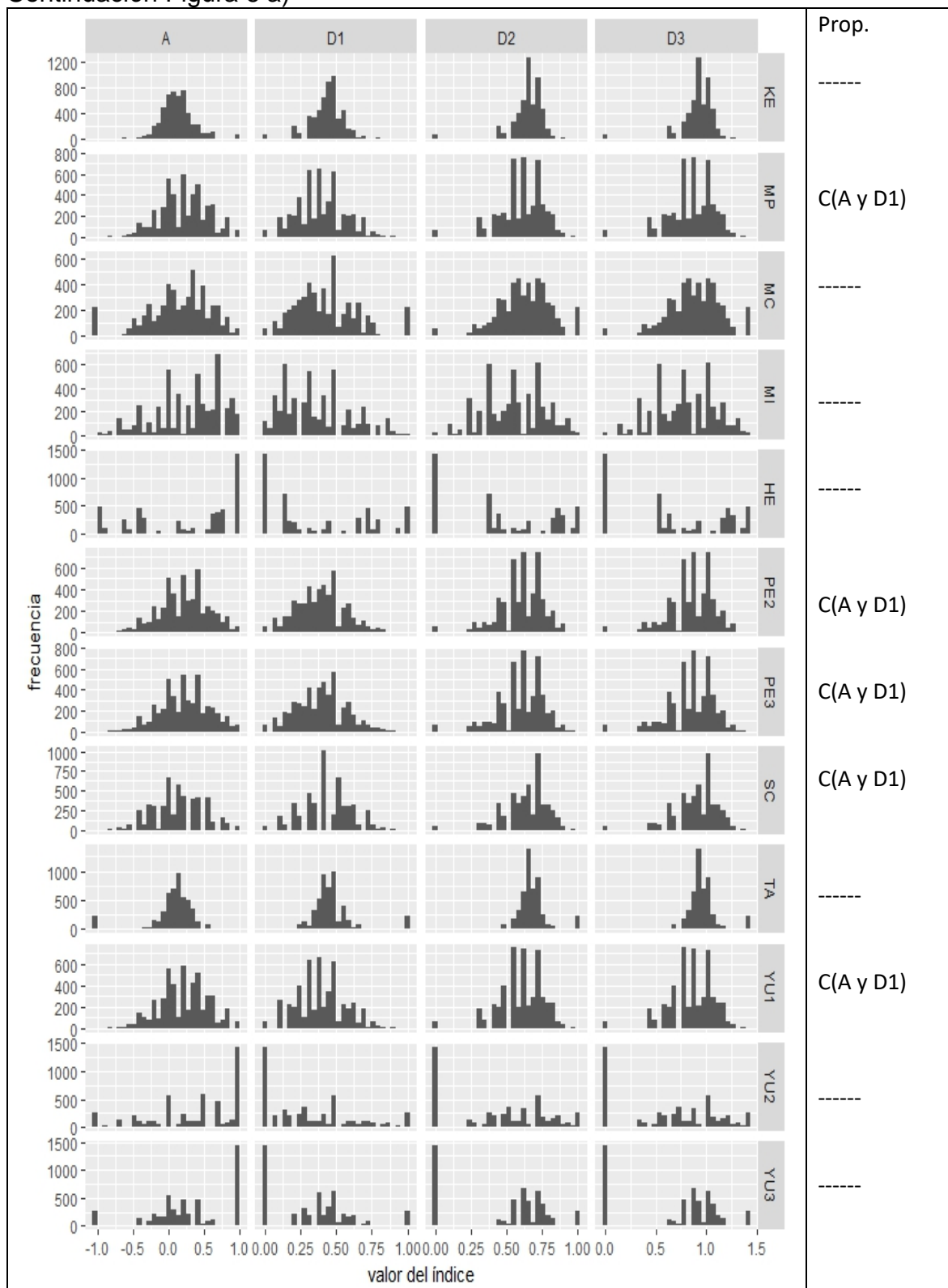
a) Índices de similitud que varían entre -1 y 1 y disimilitudes derivadas, grupo 1



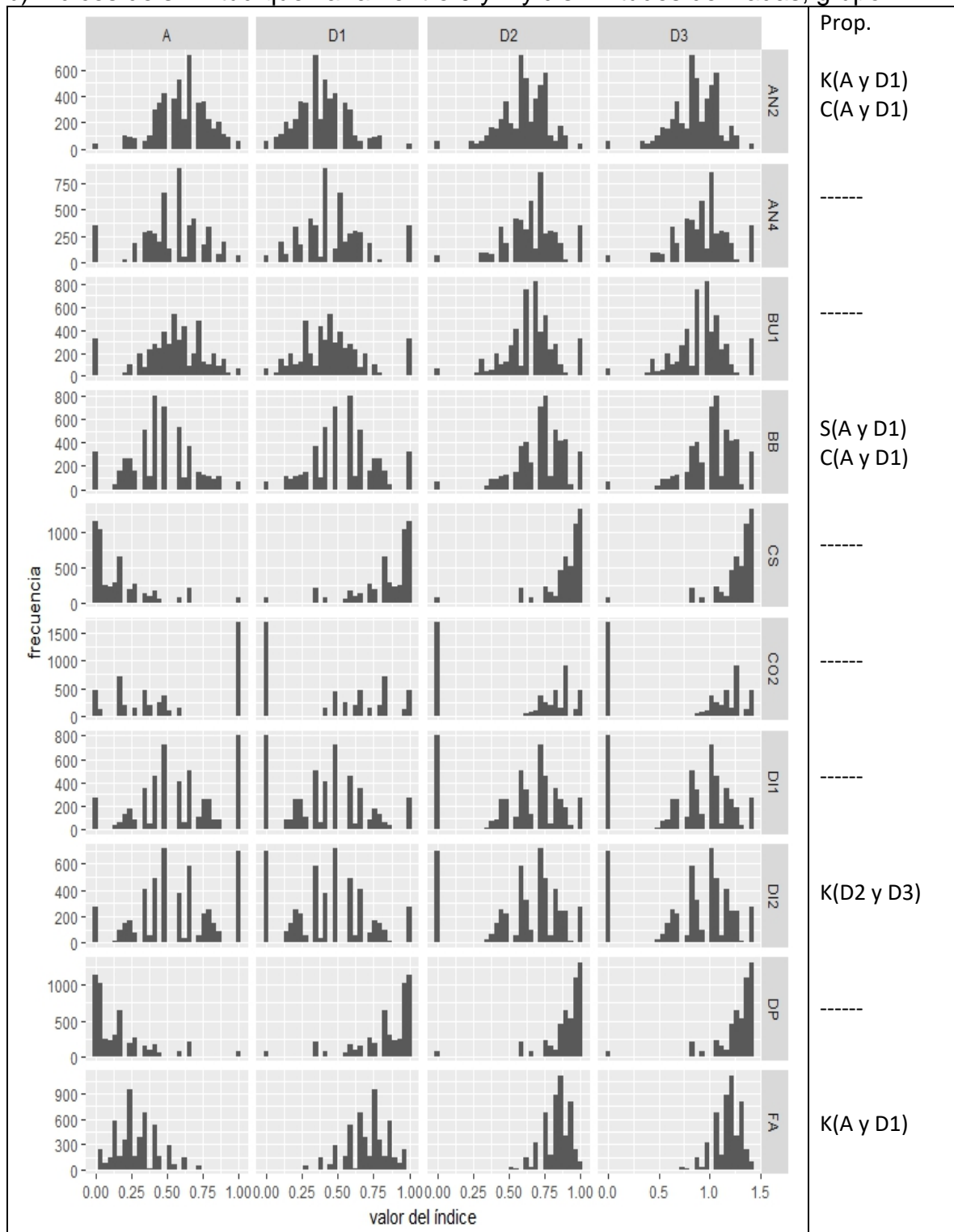
Nota 1: La columna de la derecha indica que propiedades cumplen, S(simetría), K(curtosis) y C(simetría y curtosis de una distribución normal según el gráfico de Cullen y Frey)

Nota 2: Dentro de cada grupo los índices se ordenan alfabéticamente por autor

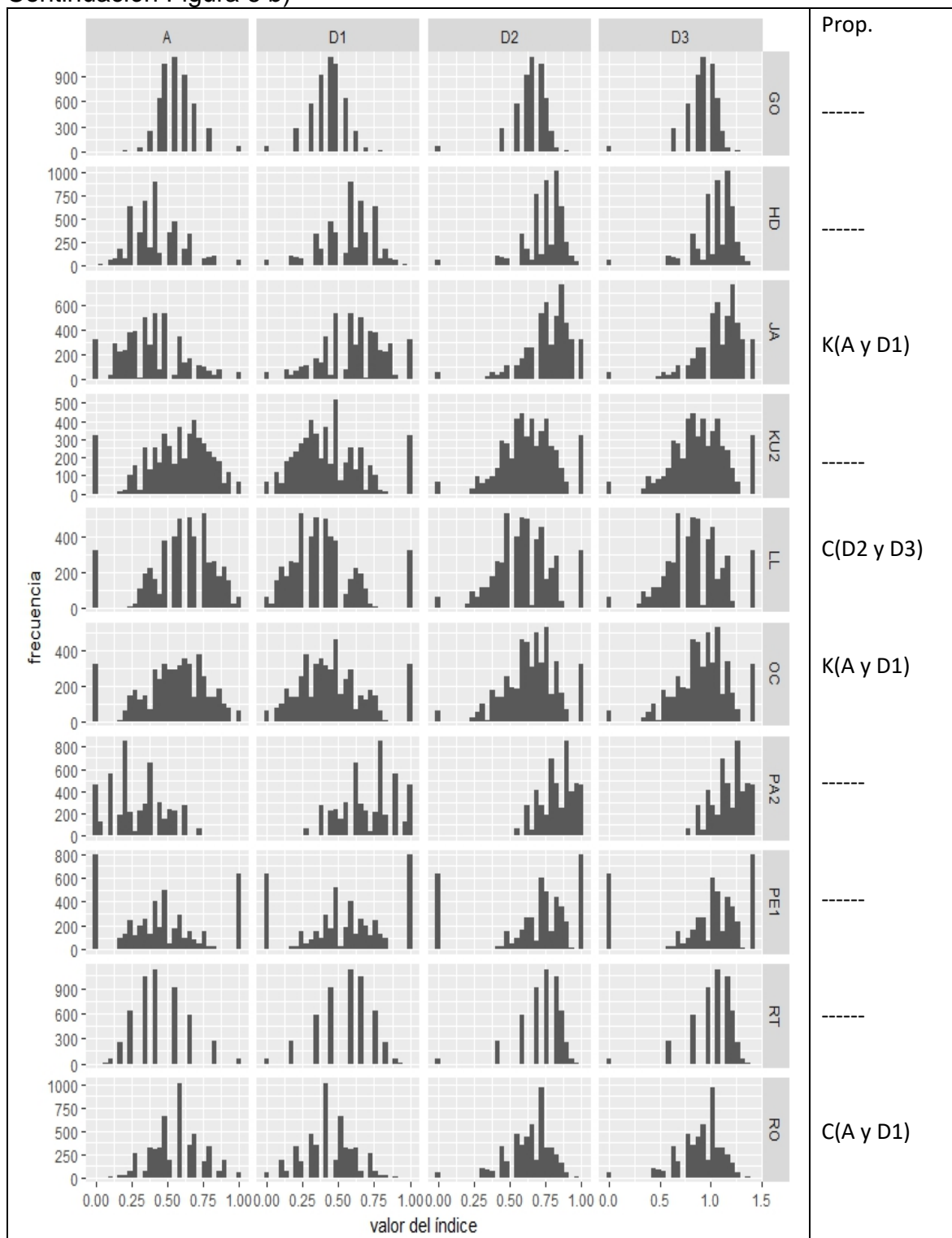
Continuación Figura 6 a)



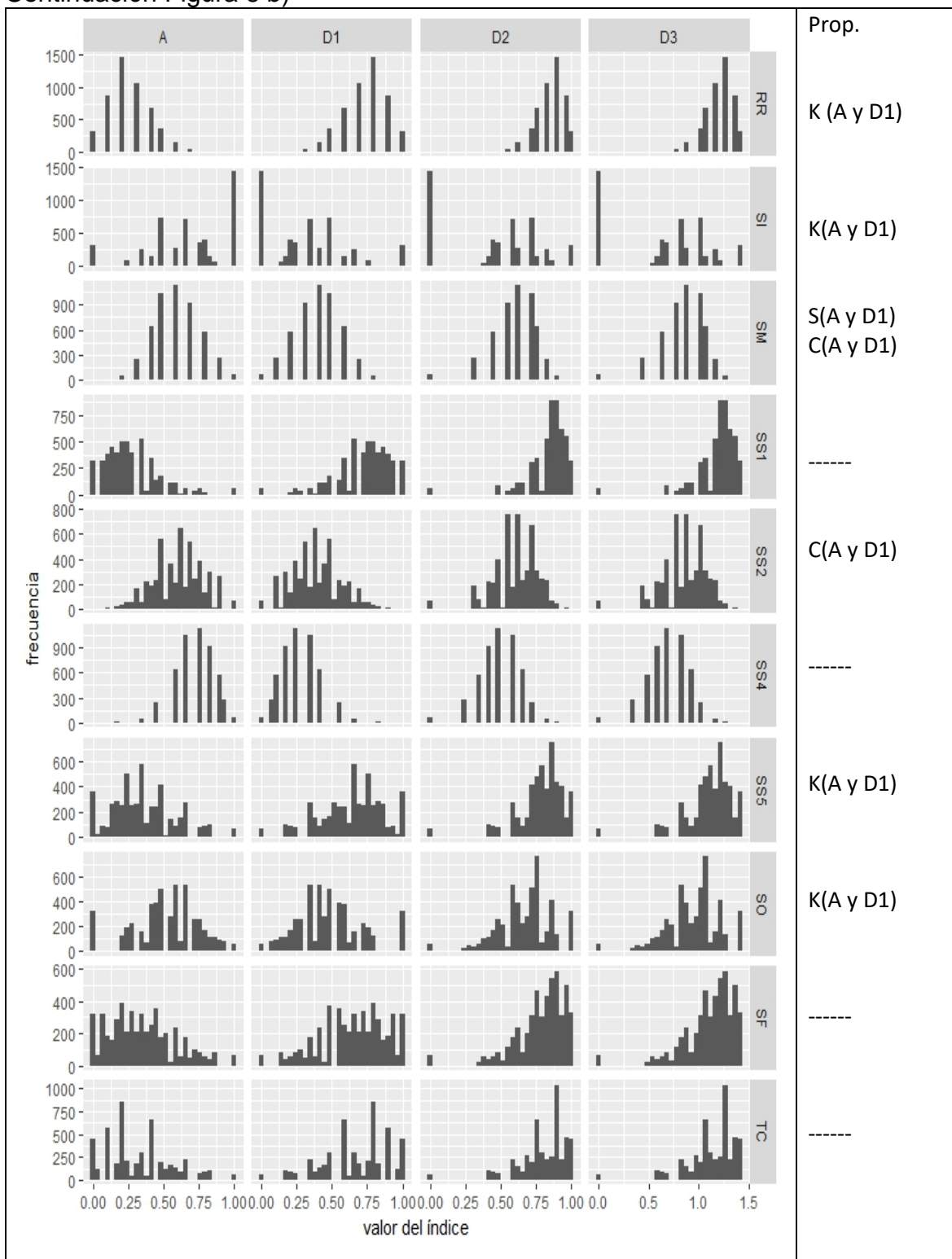
b) Índices de similitud que varían entre 0 y 1 y disimilitudes derivadas, grupo 2



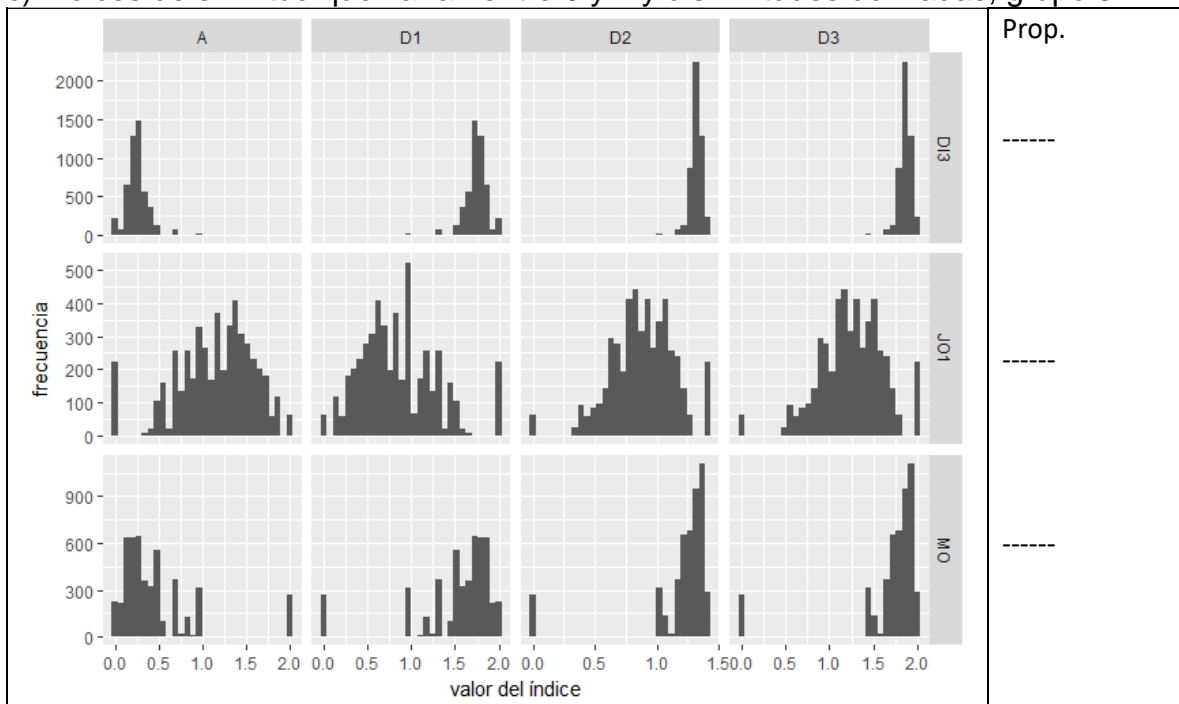
Continuación Figura 6 b)



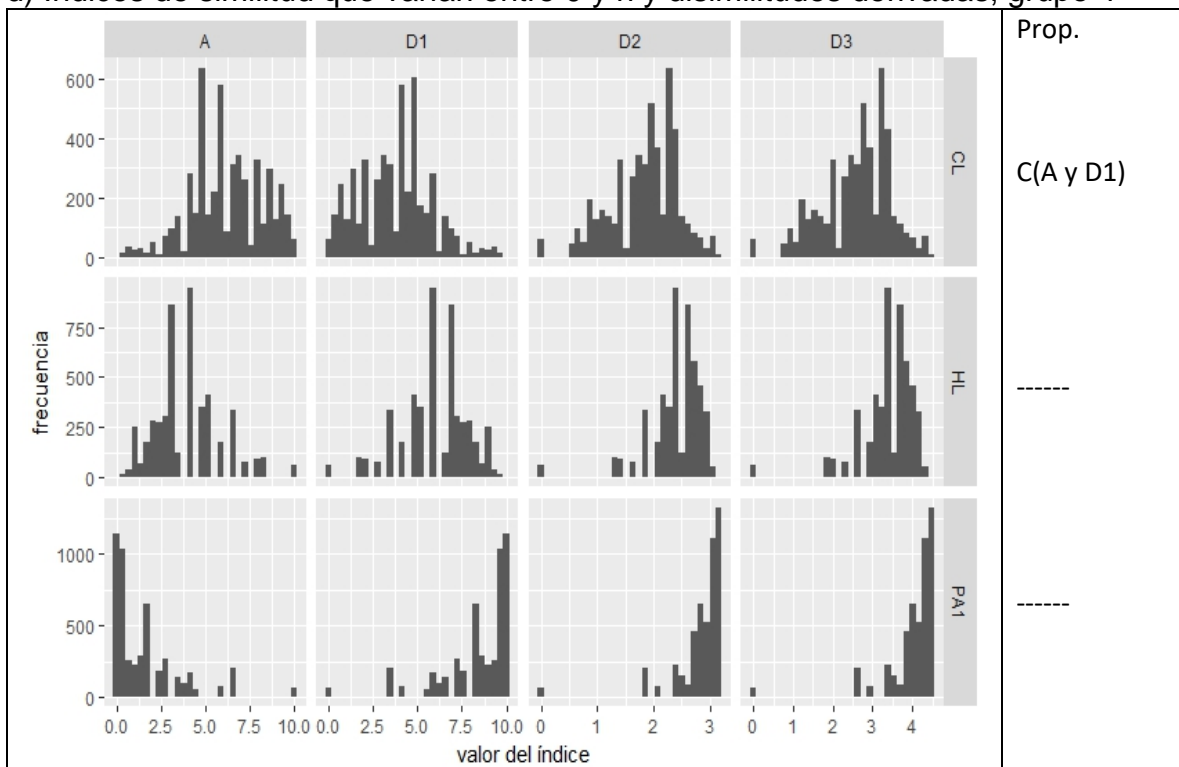
Continuación Figura 6 b)



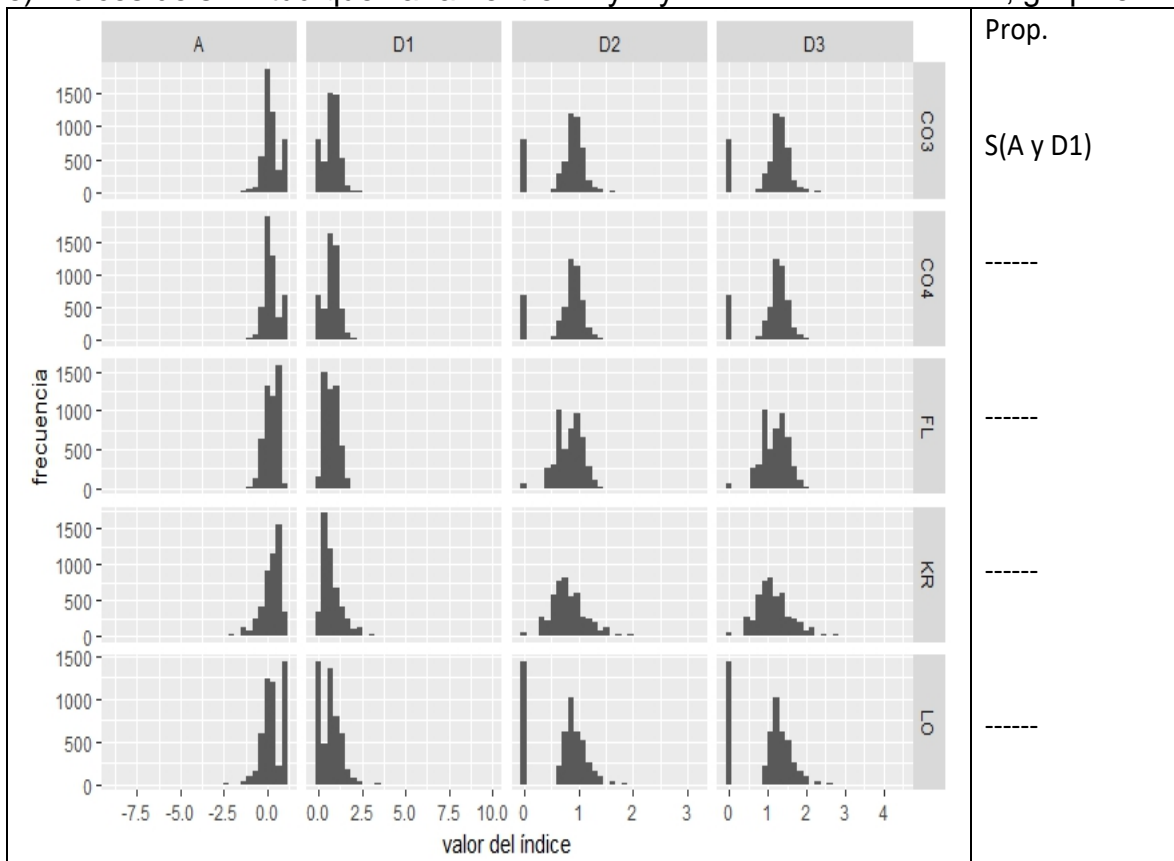
c) Índices de similitud que varían entre 0 y 2 y disimilitudes derivadas, grupo 3



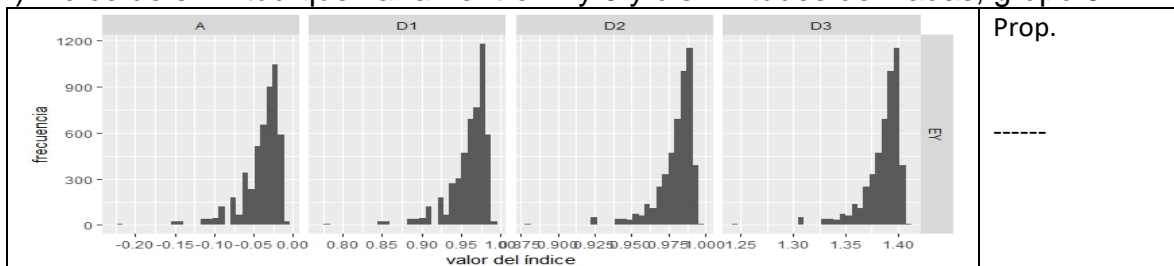
d) Índices de similitud que varían entre 0 y n y disimilitudes derivadas, grupo 4



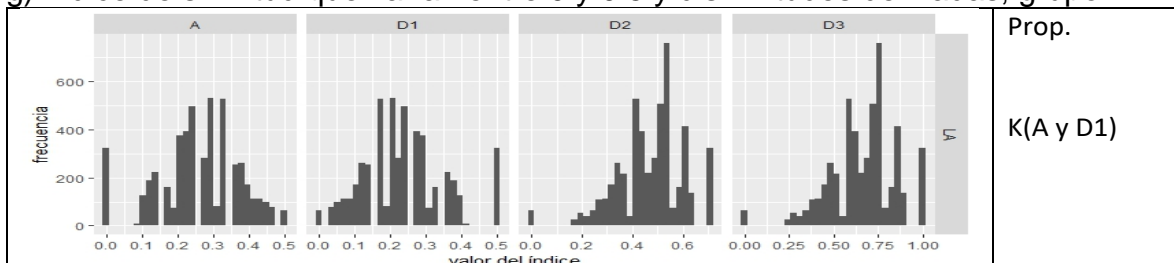
e) Índices de similitud que varían entre  $-\infty$  y 1 y disimilitudes derivadas, grupo 5



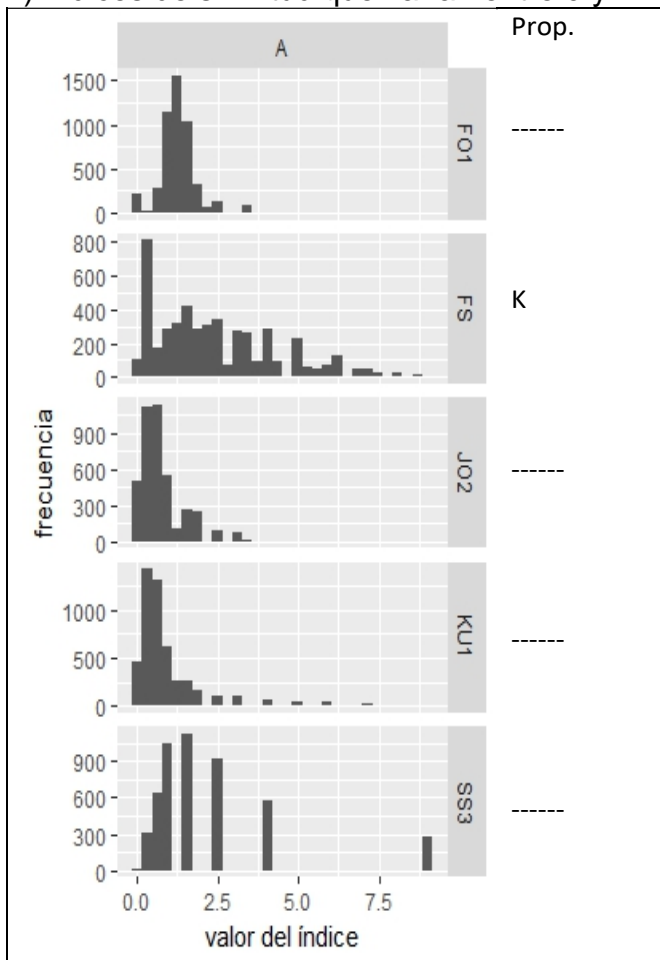
f) Índice de similitud que varían entre -1 y 0 y disimilitudes derivadas, grupo 6



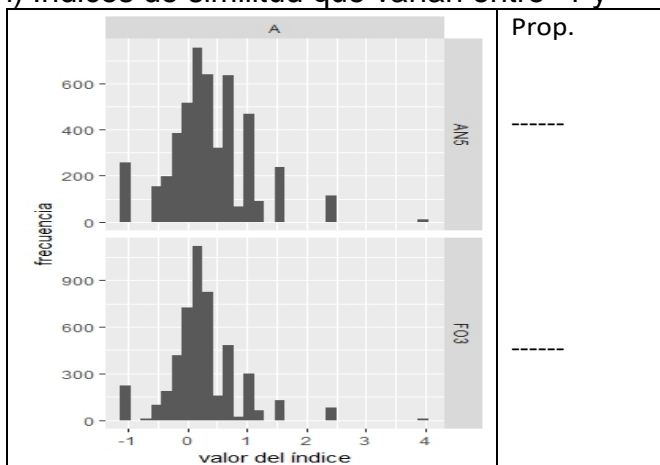
g) Índice de similitud que varían entre 0 y 0.5 y disimilitudes derivadas, grupo 7



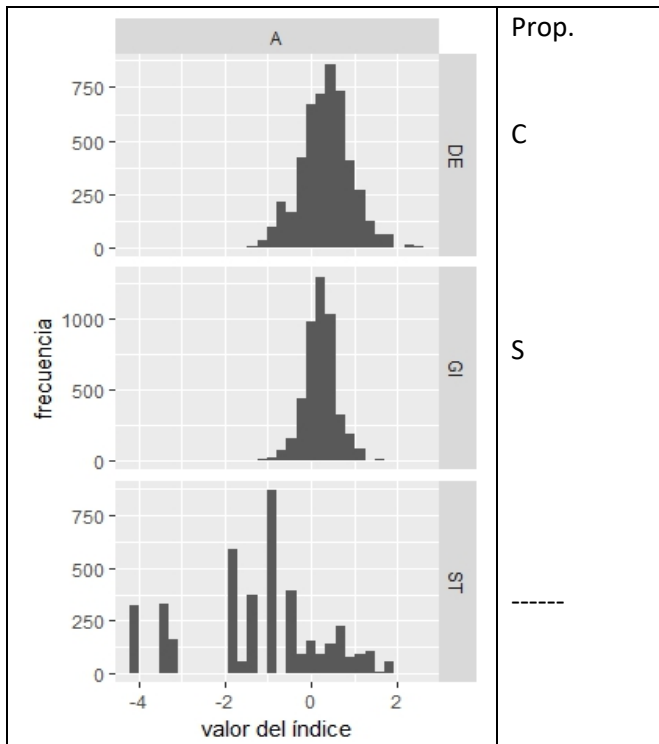
h) Índices de similitud que varían entre 0 y  $+\infty$ , grupo 8



i) Índices de similitud que varían entre -1 y  $+\infty$ , grupo 9



j) Índices de similitud que varían entre  $-\infty$  y  $+\infty$ , grupo 10



Cinco distribuciones (6%) cumplen con el supuesto de asimetría acorde a una distribución normal, las correspondientes a los índices de Hamann (grupo 1), Braun-Blanquet, Sokal & Michener (grupo 2) y Cole\_3 (grupo 5) (para  $A_i$  y  $D1_i$ ). El índice de Gilbert & Wells (grupo 10) cumple el supuesto de asimetría solamente para la similitud ( $A_i$ ) ya que no es acotado (Tabla 11 y Figura 6).

Dieciocho distribuciones (23%) cumplen el supuesto de curtosis acorde a una distribución Normal, se encuentran remarcadas en la Tabla 11 y en la Figura 6 y se corresponden a los índices de Anónimo\_1, Anónimo\_3, Cohen, Goodman & Kruskal, Hurlbert\_1, Hurlbert\_2 (grupo 1), Anónimo\_2, Faith, Jaccard, Ochiai, Russell & Rao, Simpson, Sokal & Sneath\_5, Sorensen (grupo 2) y Lamont & Grant (grupo 7) (para  $A_i$  y  $D1_i$ ). El índice de Fossum (grupo 8) cumple el supuesto de curtosis solamente para la similitud ( $A_i$ ) ya que no es acotado. Los índices de Cole\_1, Hurlbert\_1 (grupo 1) y Dice\_2 (grupo 2) cumplen el supuesto pero para las disimilitudes  $D2_i$  y  $D3_i$ .

Para veinte índices (26%), las distribuciones empíricas coincidieron con el gráfico de la distribución Normal de Cullen y Frey (1999) (Apéndice 5). Las mismas se encuentran remarcadas en la Tabla 11 y en la Figura 6 y son: Anónimo\_3, Cohen, Goodman & Kruskal, Hamann, Hurlbert\_1, Hurlbert\_2, Maxwell & Pilliner, Peirce\_2, Peirce\_3, Scott, Yule\_1 (grupo 1), Anónimo\_2, Braun-Blanquet, Rogot & Goldberg, Sokal & Michener, Sokal & Sneath\_2 (grupo 2) y Clement (grupo 4) para la similitud  $A_i$  y la disimilitud 1 ( $D1_i$ ). El índice de Dennis (grupo 10) cumple con esta propiedad solamente para la similitud ( $A_i$ ) por ser no acotado. Es de destacar que las disimilitudes 2 y 3 ( $D2_i$  y  $D3_i$ ) suelen ser más concentradas y pierden en todos los casos las propiedades de simetría y curtosis que tienen los índices de los cuales derivan. La excepción son dos índices, Fager & Mc Gowan (grupo 1) y Legendre & Legendre (grupo 2), que tienen un comportamiento opuesto.

Algunos índices presentan curtosis similar a la de la distribución normal, como Dice\_2 y Simpson (grupo 2) y Fossum (grupo 8), sin embargo sus distribuciones son bastante asimétricas (Figura 6).

Hubálek (1982) comenta en su trabajo los resultados encontrados por Baroni-Urbani & Buser en 1976, trabajando con una muestra de 39.916.800 unidades muestrales correspondiente a todas las permutaciones posibles de una unidad muestral con once variables de presencia/ausencia. Los autores encuentran que el índice de ellos, Baroni-Urbani\_1, tiene una distribución cercana a la simetría, mientras que el índice de Yule\_1 muestra una pequeña proporción de valores iguales a cero mientras que el resto de los valores calculados se encuentran concentrados entre 0.5 y 1, cuando su rango de variación es [0,1]. En este trabajo, se simulan distintas situaciones encontrando que el índice de Baroni-Urbani\_1 no cumple ninguna de las dos condiciones de asimetría y curtosis similares a la distribución normal. Por su parte para la distribución empírica del índice de Yule\_1 la asimetría/curtosis se superpone a la asimetría y curtosis de una distribución Normal. El índice de Yule\_2, resulta simétricamente distribuido en el trabajo de Baroni-Urbani & Buser, pero con una alta proporción de valores en los extremos izquierdo y derecho del rango y en este trabajo tampoco cumple con las propiedades esperadas. En la práctica el uso de este coeficiente para cálculos sobre distintas matrices de datos bien conocidas ha mostrado una

supervaloración exagerada de la similitud, con una alta proporción de valores en los extremos del rango de variación (Baroni-Urbani & Buser 1976).

### **Propiedad 12: Insesgamiento**

Para probar esta propiedad se construyeron intervalos de confianza para los valores poblacionales de cada índice considerando repeticiones de distintos escenarios y distintos tamaños de muestra. Se contabilizan las situaciones en que los mismos contienen al valor poblacional (cobertura). Se calculan los sesgos y SE.

Más de la mitad de los índices (69%) son insesgados, encontrándose que en todos los casos de insesgamiento, los intervalos de confianza cubren al verdadero valor poblacional (cobertura del 100%). El 29% de los índices tienen una cobertura de entre el 65% y el 95%. Solamente el índice de Hurlbert\_1 tiene un 50% de cobertura y el índice de Pearson & Heron apenas un 5%. La Tabla 12 muestra los resultados encontrados para cada índice de similitud, agrupados de acuerdo al porcentaje de cobertura, de mayor a menor. Dentro de cada grupo de igual cobertura, se ordenan los índices por magnitud del sesgo. De los índices con mayor cobertura encontramos un sesgo promedio más alto para los índices de Harris & Lahey y Clement, pero son los que presentan mayor variabilidad. Para los sesgados se observa que Pearson\_1 presenta un sesgo promedio más alto y tiene verdaderamente una SE mayor.

Los resultados de los promedios y desvíos muestrales y poblacionales de todas las simulaciones se encuentran en el Anexo 5.

Estos resultados coinciden con lo obtenido por Goodall en 1973, a partir de un número menor de repeticiones en los mismos escenarios, que observa que los índices de Jaccard y Yule\_1 son en general insesgados, aún con muestras pequeñas y que Cole\_2 es sesgado. Sin embargo para los índices de Sokal & Michener y Sorensen en este trabajo no se encuentran discrepancias significativas entre los valores muestrales y poblacionales mientras que Goodall señala que estimados desde muestras pequeñas tienden a subestimar el valor poblacional.

Tabla 12: Insegamiento de los índices de similitud para datos binarios.

Índice de similitud	Cobertura en %	Sesgo promedio	SE promedio
Eyraud	100	0,000000	0,000004
Goodall	100	0,000047	0,006148
Rogot & Goldberg	100	0,000086	0,008019
Sokal & Michener	100	0,000094	0,008802
Rogers & Tanimoto	100	0,000107	0,009128
Ochiai	100	0,000203	0,012075
Fager & Mc Gowan	100	0,000207	0,012270
Legendre & Legendre	100	0,000249	0,012896
Peirce_2	100	0,000260	0,014250
Anónimo_4	100	0,000280	0,013251
Cohen	100	0,000291	0,014480
Maxwell & Pilliner	100	0,000314	0,015602
Anónimo_3	100	0,000337	0,016104
Yule_1	100	0,000344	0,016596
Scott	100	0,000346	0,016053
Hamann	100	0,000377	0,017590
Cole_1	100	0,000521	0,019555
Sokal & Sneath_2	100	0,000594	0,009148
Anónimo_1	100	0,000756	0,022809
Dice_1	100	0,000782	0,020816
Peirce_3	100	0,000881	0,024052
Mc Connaughey	100	0,001185	0,028293
Johnson_1	100	0,001207	0,028330
Cole_3	100	0,002529	0,036739
Kuder & Richardson	100	0,002623	0,037338
Yule_2	100	0,002744	0,036254
Digby	100	0,002865	0,037693
Cole_4	100	0,003209	0,014452
Fleiss	100	0,003335	0,045115
Kendall	100	0,004038	0,009471
Dice_2	100	0,004458	0,010576
Anónimo_2	100	0,004644	0,008062
Sorensen	100	0,005024	0,011387
Kulczynski_2	100	0,005182	0,014167
Sokal & Sneath_5	100	0,005475	0,010274
Sokal & Sneath_4	100	0,006517	0,007306
Forbes_1	100	0,006921	0,060907
Anónimo_5	100	0,007033	0,061112
Forbes_3	100	0,007165	0,060970
Jaccard	100	0,008195	0,009121
Peirce_1	100	0,008581	0,008733
Hawkins & Dotson	100	0,008728	0,007876
Kulczynski_1	100	0,008887	0,030333
Lamont & Grant	100	0,011961	0,005701
Dice_3	100	0,014231	0,001218
Faith	100	0,014626	0,005532
Sokal & Sneath_1	100	0,016320	0,006844
Dennis	100	0,016947	0,114671
Russell & Rao	100	0,020080	0,004662
Sokal & Sneath_3	100	0,044916	0,108317
Harris & Lahey	100	1,097617	0,817281
Clement	100	6,664688	1,877815
Michael	95	0,000625	0,019719
Tarwid	95	0,004017	0,031693
Mountford	90	0,009338	0,002436
Baroni-Urbani & Buser_1	85	0,000431	0,013185
Simpson_1	85	0,000996	0,020559
Baroni-Urbani & Buser_2	85	0,001702	0,026393
Sorgenfrei	85	0,003148	0,008479
Braun-Blanquet	85	0,004993	0,010081
Yule_3	85	0,005045	0,039698
Johnson_2	85	0,091855	0,118023
Clifford & Stephenson	80	0,000245	0,011028
Doolittle	80	0,000247	0,011044
Pearson_2	80	0,000751	0,012282
Tschuproff	80	0,000852	0,015042
Cole_2	80	0,016054	0,035872
Gilbert & Wells	80	0,032591	0,061548
Hurlbert_2	80	0,482567	0,132900
Pearson_1	80	2,424093	1,104652
Goodman & Kruskal	75	0,001093	0,021700
Fossum	75	0,100627	0,213052
Stiles	75	0,777883	0,172514
Loevinger	65	0,003513	0,036090
Forbes_2	65	0,003908	0,012748
Hurlbert_1	50	0,014792	0,029347
Pearson & Heron	5	0,401532	0,135719

Nota: el sombreado corresponde a valores altos del sesgo promedio

### **Propiedad 13: Métrica**

Se presentan los resultados de la prueba de la desigualdad triangular en los tres índices de disimilitud derivados de los setenta y siete índices de similitud estudiados en este trabajo (Tabla 13). Se trabaja con la matriz de datos binarios simulada.

Para la Disimilitud 1 ( $D1=1-A$ ) solamente un 18% de los índices cumplen esta propiedad, mientras que más de la mitad (59%) no la cumplen y el restante 23% corresponde a índices de similitud que no son simétricos, condición necesaria junto con la positividad para una métrica, o corresponden a índices de similitud no acotados que por lo tanto no permiten derivar las disimilitudes.

Para las disimilitudes 2 y 3 ( $D2 = \sqrt{(1 - A)}$  y  $D3 = \sqrt{2(1 - A)}$ ) casi la mitad de los índices (47%) cumplen la propiedad métrica. El 30% no la cumplen y el 23% restante no permite calcularla.

El 18% de los índices cumplen esta propiedad para las tres disimilitudes estudiadas. Ellos son: Braun-Blanquet, Dice\_3, Eyraud, Faith, Forbes\_2, Goodall, Hamann, Hawkins & Dotson, Jaccard, Kendall, Roger & Tanimoto, Russell & Rao, Sokal & Michener y Sokal & Sneath\_1 (Tabla 13).

Estos resultados coinciden con lo obtenido por Gower & Legendre (1986) y Cuadras (1989) para los catorce índices analizados por ellos, estudiando la Disimilitud 1 y la Disimilitud 2. Ellos encontraron seis índices que cumplen la propiedad métrica para ambas disimilitudes: Hamann, Jaccard, Rogers & Tanimoto, Russell & Rao, Sokal & Michener y Sokal & Sneath\_1. Cinco índices la cumplen solo para la Disimilitud 2: Ochiai, Sokal & Sneath\_4, Sokal & Sneath\_5, Sorensen y Yule\_1, mientras que el resto no la cumplen para ninguna de las dos disimilitudes estudiadas: Kulczynski\_2, Sokal & Sneath\_2 y Yule\_2.

En el caso del índice de Sokal & Sneath\_2 según los resultados obtenidos, no se cumple la desigualdad triangular para la Disimilitud 1, sin embargo, si se cumple para las Disimilitudes 2 y 3, pero en el trabajo de Gower & Legendre (1986) se presenta un contra ejemplo donde se observa que no se verifica la propiedad, por lo tanto se considera que el índice es no métrico para las 3 Disimilitudes (cálculos en el Anexo 6).

Los resultados de esta tesis coinciden con los obtenidos por Giansanti (1990) para veintitres índices, con la excepción de que ella considera asimétricos a los índices de Braun-Blanquet y Simpson, los que según las definiciones adoptadas son simétricos y por lo tanto es posible estudiar sus propiedades métricas. Estos resultados son avalados por Demey et al. (2011), quienes presentan las propiedades métricas de un conjunto más amplio de catorce índices, que incluyen a Braun-Blanquet y Simpson. Ellos encuentran que ambos índices son métricos para las Disimilitudes 2 y 3, lo cual corrobora que son simétricos. Sin embargo, el índice de Simpson no resulta métrico en este trabajo, coincidiendo con Bren & Batagelj (2006). Estos autores estudian las propiedades métricas de dieciséis índices de disimilitud (D1 y D2), incluidos en este trabajo, y sus resultados coinciden completamente con lo encontrado aquí. Esta discrepancia entre autores surge de considerar para el estudio de la propiedad métrica solo algún ejemplo. Puede cumplirse para un caso particular y sin embargo no generalizarse para otros.

Warrens (2008) trabajando con familias de coeficientes llega a los mismos resultados que en este trabajo, pero solo trabajando con la Disimilitud 1.

Tabla 13: Propiedad Métrica de los Índices de disimilitud para datos binarios (D1, D2 y D3)

Grupos	Indice de similitud	D1	D2	D3
	Braun-Blanquet	si	si	si
	Dice_3	si	si	si
	Eyraud	si	si	si
	Faith	si	si	si
	Forbes_2	si	si	si
<b>Tres disimilitudes métricas</b>	Goodall	si	si	si
	Hamann	si	si	si
	Hawkins & Dotson	si	si	si
	Jaccard	si	si	si
	Kendall	si	si	si
	Rogers & Tanimoto	si	si	si
	Russell & Rao	si	si	si
	Sokal & Michener	si	si	si
	Sokal & Sneath_1	si	si	si
		Anónimo_1	no	si
	Anónimo_2	no	si	si
	Anónimo_3	no	si	si
	Clifford & Stephenson	no	si	si
	Cohen	no	si	si
	Doolittle	no	si	si
	Fager & Mc Gowan	no	si	si
	Goodman & Kruskal	no	si	si
	Harris & Lahey	no	si	si
<b>Dos disimilitudes métricas</b>	Lamont & Grant	no	si	si
	Legendre & Legendre	no	si	si
	Maxwell & Pilliner	no	si	si
	Ochiai	no	si	si
	Pearson_1	no	si	si
	Rogot & Goldberg	no	si	si
	Scott	no	si	si
	Sokal & Sneath_4	no	si	si
	Sokal & Sneath_5	no	si	si
	Sorensen	no	si	si
	Sorgenfrei	no	si	si
	Tschuproff	no	si	si
	Yule_1	no	si	si
		Anónimo_4	no	no
	Baroni-Urbani & Buser_1	no	no	no
	Baroni-Urbani & Buser_2	no	no	no
	Cole_1	no	no	no
	Cole_2	no	no	no
	Digby	no	no	no
	Fleiss	no	no	no
	Hurlbert_1	no	no	no
	Hurlbert_2	no	no	no
	Johnson_1	no	no	no
<b>Ninguna disimilitud métrica</b>	Kuder & Richardson	no	no	no
	Kulczynski_2	no	no	no
	Loevinger	no	no	no
	Mc Connaughey	no	no	no
	Michael	no	no	no
	Mountford	no	no	no
	Pearson_2	no	no	no
	Pearson & Heron	no	no	no
	Simpson	no	no	no
	Sokal & Sneath_2	no	no	no
	Tarwid	no	no	no
	Yule_2	no	no	no
	Yule_3	no	no	no
	Clements	*	*	*
	Cole_3	*	*	*
	Cole_4	*	*	*
	Dice_1	*	*	*
	Dice_2	*	*	*
	Peirce_1	*	*	*
	Peirce_2	*	*	*
	Peirce_3	*	*	*
<b>Disimilitudes no evaluables</b>	Anónimo_5	#	#	#
	Dennis	#	#	#
	Forbes_1	#	#	#
	Forbes_3	#	#	#
	Fossum	#	#	#
	Gilbert & Wells	#	#	#
	Kulczynski_1	#	#	#
	Sokal & Sneath_3	#	#	#
	Stiles	#	#	#
	Johnson_2	@	@	@

Nota: el símbolo \* significa que el índice de disimilitud es no acotado, el # no simétrico y el @ ambas cosas, por lo cual no se puede probar la propiedad métrica.

#### **Propiedad 14: Indeterminancia**

El 36% de los índices de similitud, sobre un total de setenta y siete, no resulta indeterminado en ninguna de las situaciones consideradas, que resultan de comparar de a pares a 100 vectores simulados, resultando 4950 casos. El 64% resultan indeterminados en alguna instancia (Tabla 14). Toda vez que la similitud (A) sea indeterminada lo son las disimilitudes derivadas de ella (D1, D2 y D3). El índice que resulta con mayor número de indeterminancias es Mountford. Para diez índices (13%) no acotados, se analizaron las indeterminancias solamente en la similitud, y resultaron todos indeterminados: Kulczynski\_1, Sokal & Sneath\_3, Anónimo\_5, Dennis, Fossum, Forbes\_1, Forbes\_3, Gilbert & Wells, Stiles y Johnson\_2. La mayor frecuencia de indeterminaciones se encuentra para los tres últimos. Cabe destacar que el índice de Stiles resulta indeterminado en 99 casos y no definido en 719 casos, ya que se trata de un logaritmo y por lo tanto no está definido para el valor de cero.

Estos resultados son similares a los del estudio de Warrens (2008), quien calcula el número de veces que veintiocho índices de similitud se hacen indeterminados en función del número de variables binarias ( $n$ ). Los índices que nunca se hacen indeterminados según Warrens son: Hamann, Michael, Rogers & Tanimoto, Russell & Rao y Sokal & Michener, los cuales en este trabajo tampoco registran indeterminancias. Para los cinco índices que Warrens (2008) considera con una indeterminancia de uno (Baroni-Urbani & Buser\_1, Braun-Blanquet, Jaccard, Sokal & Sneath\_1 y Sorensen), en este trabajo no se obtuvieron indeterminancias y es más, se considera que no son indeterminados, ya que la indeterminancia se produce cuando se comparan dos unidades muestrales con todos sus valores iguales a cero. Para los casos de dos indeterminancias (Cohen, Goodman & Kruskal, Hawkins & Dotson y Scott) y cuatro (Maxwell & Pilliner) según Warrens, no se obtuvieron ninguna en este trabajo. Si se encuentran indeterminancias para los casos de  $2n+1$  (Kulczynski\_2, Mc Connaughey, Ochiai, Simpson y Sorgenfrei) y  $4n$  (Loevinger, Sokal & Sneath\_5, Yule\_2 y Yule\_3). Las únicas diferencias son con respecto a Sokal & Sneath\_2, para el cual nosotros encontramos 99 indeterminancias mientras que para Warrens nunca se hace indeterminado y para Sokal & Sneath\_4, que no resulta indeterminado en esta tesis, mientras que Warrens reporta un número de

indeterminancias de  $4n$ . Consideramos que existe un error en los números de los coeficientes, ya que este último índice tiene en el denominador la expresión  $2a+b+c+2d$ , la cual sería cero solamente si las cuatro cantidades  $a$ ,  $b$ ,  $c$  y  $d$  fueran todas iguales a cero (ver fórmulas de los índices en página 32).

Tabla 14: Propiedad de Indeterminancia de los índices de similitud (A) y disimilitud (D1, D2 y D3).

Grupos	Índice de similitud	Número de indeterminancias				
		A	D1	D2	D3	
	Anónimo_1	-	-	-	-	
	Anónimo_2	-	-	-	-	
	Anónimo_3	-	-	-	-	
	Baroni-Urbani & Buser_1	-	-	-	-	
	Baroni-Urbani & Buser_2	-	-	-	-	
	Braun-Blanquet	-	-	-	-	
	Cohen	-	-	-	-	
	Faith	-	-	-	-	
	Forbes_2	-	-	-	-	
	Goodall	-	-	-	-	
	Goodman & Kruskal	-	-	-	-	
	Hamann	-	-	-	-	
<b>No presentan indeterminancias</b>	Harris & Lahey	-	-	-	-	
	Hawkins & Dotson	-	-	-	-	
	Jaccard	-	-	-	-	
	Kendall	-	-	-	-	
	Lamont & Grant	-	-	-	-	
	Legendre & Legendre	-	-	-	-	
	Maxwell & Pilliner	-	-	-	-	
	Michael	-	-	-	-	
	Rogers & Tanimoto	-	-	-	-	
	Rogot & Goldberg	-	-	-	-	
	Russell & Rao	-	-	-	-	
	Scott	-	-	-	-	
	Sokal & Michener	-	-	-	-	
	Sokal & Sneath_1	-	-	-	-	
	<b>Sokal &amp; Sneath_4</b>	-	-	-	-	
	Sorensen	-	-	-	-	
		Anónimo_4	1	1	1	1
		Kuder & Richardson	1	1	1	1
		Cole_4	47	47	47	47
		Dice_2	47	47	47	47
	Peirce_2	47	47	47	47	
	Clement	52	52	52	52	
	Cole_3	52	52	52	52	
	Dice_1	52	52	52	52	
	Peirce_3	52	52	52	52	
	Cole_1	99	99	99	99	
	Cole_2	99	99	99	99	
	Dice_3	99	99	99	99	
	Digby	99	99	99	99	
	Doolittle	99	99	99	99	
	Hurlbert_1	99	99	99	99	
	Clifford & Stephenson	99	99	99	99	
	Eyraud	99	99	99	99	
	Fager & Mc Gowan	99	99	99	99	
<b>Presentan indeterminancias</b>	Fleiss	99	99	99	99	
	Johnson_1	99	99	99	99	
	Kulczynski_2	99	99	99	99	
	Loevinger	99	99	99	99	
	Mc Connaughey	99	99	99	99	
	Ochiai	99	99	99	99	
	Pearson & Heron	99	99	99	99	
	Pearson_1	99	99	99	99	
	Pearson_2	99	99	99	99	
	Simpson	99	99	99	99	
	<b>Sokal &amp; Sneath_2</b>	99	99	99	99	
	Sokal & Sneath_5	99	99	99	99	
	Sorgenfrei	99	99	99	99	
	Tarwid	99	99	99	99	
	Tschuproff	99	99	99	99	
	Yule_1	99	99	99	99	
	Yule_2	99	99	99	99	
	Yule_3	99	99	99	99	
	Peirce_1	110	110	110	110	
	Hurlbert_2	147	147	147	147	
Mountford	162	162	162	162		
	Kulczynski_1	63	#	#	#	
	Sokal & Sneath_3	63	#	#	#	
<b>No acotados con indeterminancias</b>	Anónimo_5	99	#	#	#	
	Dennis	99	#	#	#	
	Fossum	99	#	#	#	
	Forbes_1	99	#	#	#	
	Forbes_3	99	#	#	#	
	Gilbert & Wells	323	#	#	#	
	Stiles	818	#	#	#	
	Johnson_2	853	#	#	#	

Nota: el símbolo # corresponde a índices de similitud no acotados no permitiendo el cálculo de las disimilitudes. El sombreado indica valores altos de indeterminancia

### Propiedad 15: Euclídeanidad

Para el estudio de la euclídeanidad se realiza un análisis a través de los autovalores (*eigen analysis*). El número de autovalores negativos encontrados para la matriz de similitud así como para las tres matrices de disimilitud de cada uno de los setenta y siete índices, se muestra en la Tabla 15. Para los casos en que se encuentran autovalores negativos se indica, entre paréntesis, el sesgo introducido en la inercia explicada para lograr una representación euclídeana desde una medida que no lo es. Si para una matriz de similitud ( $A_{rxr}$ ) no se obtienen autovalores negativos eso implica que la misma es definida positiva o semi definida positiva (psd). Si no se obtienen autovalores negativos para alguna o todas las matrices de disimilitud estudiadas ( $D1_{rxr}$ ,  $D2_{rxr}$  y  $D3_{rxr}$ ) eso indica que las mismas son euclídeanas.

Un 39% de los índices definieron matrices psd para la similitud y euclídeanas para las disimilitudes 2 y 3. Un 36% de los índices no produjeron matrices psd ni euclídeanas para ninguna de las medidas estudiadas. Existen doce índices (16%) que no permitieron estudiar esta propiedad, por no ser métricos, simétricos, o hacerse indeterminados en la diagonal principal. Los índices de Clement, Cole\_3, Cole\_4, Dice\_1, Dice\_2, Peirce\_2 y Peirce\_3 son no simétricos, y esta es una condición previa a la Euclídeanidad. El índice de Mountford resulta indeterminado en la diagonal principal y por lo tanto no permite realizar el análisis de autovalores. El índice de Johnson\_2 es no simétrico, no acotado e indeterminado por lo tanto tampoco permitió realizar el análisis de autovalores. El índice de Peirce\_1 es no simétrico e indeterminado. Los índices de Kulczynski\_1 y Sokal & Sneath\_3 también producen matrices indeterminadas en la diagonal principal y además son no acotados. Para un conjunto de siete índices (9%) no acotados: Dennis, Forbes\_1, Forbes\_3, Fossum, Anónimo\_5, Gilbert & Wells y Stiles, que solamente permiten evaluar si la matriz de similitud es psd, solo los tres últimos no cumplen esta propiedad, el resto si (Tabla 15).

Para la matriz de disimilitud 1 ( $D1=1-A$ ) no se encontró ningún resultado favorable. Legendre & Legendre (1998) habían reportado que la mayoría de las disimilitudes que se derivan desde la transformación de un coeficiente de similitud, utilizando la fórmula  $D=1-A$  no son euclídeas.

En aquellos índices de similitud y/o disimilitud para los cuales se encuentran autovalores negativos solo para la disimilitud 1, se destacan los valores bajos del sesgo introducido (representado para valores negativos de la inercia) correspondiente a Dice\_3 y Kendall, mientras que para Ochiai el valor es alto. Para los índices que producen autovalores negativos para las cuatro medidas estudiadas (A, D1, D2 y D3) se encuentran valores bajos de sesgo para Fager & Mc Gowan, Kulczynski\_2 y Mc Connaughey, mientras que para Kuder & Richardson, Pearson & Heron y Yule\_2 los valores son altos. Para los índices no acotados se destaca el valor bajo de sesgo para Anonimo\_5 (Tabla 15). Debe notarse que los porcentajes de sesgo encontrados para algunos índices superan, en varios casos el 100%, indicando que el sesgo introducido para la representación euclídeana, en particular con un número reducido de dimensiones, difícilmente pueda reflejar las relaciones existentes en la nube de puntos, sino más bien el sesgo mismo.

Nuestros resultados coinciden completamente con lo obtenido por Gower & Legendre (1986) para catorce índices y Giansanti (1990) para veintitrés índices. Esta última no realizó los cálculos para los índices de Braun-Blanquet y Simpson por considerarlos erróneamente asimétricos. Sin embargo como se expresa en la Propiedad 13 (Métrica) estos resultados son avalados por Demey et al. (2011), quienes presentan la propiedad psd para la similitud y las propiedades euclídeanas de catorce índices de disimilitud, basándose en una tabla de Cuadras (1996). Sus resultados corresponden a las propiedades euclídeanas de las Disimilitudes 2 y 3, y los mismos coinciden con lo obtenido en este trabajo para el índice de Braun-Blanquet entre otros. Por el contrario, el índice de Simpson cumple la propiedad para Demey et al. (2011), sin embargo, aquí se muestra que la matriz de similitud no es psd y que las disimilitudes 2 y 3 no resultan euclídeanas.

Tabla 15: Cantidad de autovalores negativos y porcentaje de sesgo encontrado en la inercia explicada en las matrices de similitud (A) y disimilitud (D1, D2 y D3)

Indice de similitud	A	D1	D2	D3
Anónimo_1	0	55 (60%)	0	0
Anónimo_2	0	55 (62%)	0	0
Anónimo_3	0	55 (63%)	0	0
Braun-Blanquet	0	48 (38%)	0	0
Clifford&Stephenson	0	36 (27%)	0	0
Cohen	0	48 (54%)	0	0
Dice_3	0	44 (9%)	0	0
Doolittle	0	36 (27%)	0	0
Faith	0	44 (25%)	0	0
Forbes 2	0	38 (38%)	0	0
Goodall	0	37 (18%)	0	0
Hamann	0	44 (63%)	0	0
Hawkins & Dotson	0	53 (21%)	0	0
Harris & Lahey	0	48 (19%)	0	0
Jaccard	0	53 (20%)	0	0
Kendall	0	27 (8%)	0	0
Lamont & Grant	0	55 (60%)	0	0
Maxwell & Pilliner	0	51 (66%)	0	0
Ochiai	0	50 (69%)	0	0
Pearson_1	0	36 (27%)	0	0
Rogers & Tanimoto	0	55 (27%)	0	0
Rogot & Goldberg	0	55 (56%)	0	0
Russell & Rao	0	44 (13%)	0	0
Scott	0	55 (56%)	0	0
Sokal & Michener	0	44 (63%)	0	0
Sokal & Sneath_1	0	55 (60%)	0	0
Sokal & Sneath_5	0	49 (28%)	0	0
Sorensen	0	55 (60%)	0	0
Sorgenfrei	0	46 (27%)	0	0
Yule_1	0	44 (67%)	0	0
Dennis	0	#	#	#
Forbes_1	0	#	#	#
Forbes_3	0	#	#	#
Fossum	0	#	#	#
Anónimo_4	28 (9%)	41 (86%)	28 (9%)	28 (9%)
Baroni-Urbani & Buser_1	28 (7%)	43 (76%)	28 (7%)	28 (7%)
Baroni-Urbani & Buser_2	28 (7%)	43 (76%)	28 (7%)	28 (7%)
Cole_1	45 (26%)	37 (114%)	45 (26%)	45 (26%)
Cole_2	35 (123%)	34 (181%)	35 (123%)	35 (123%)
Digby	53 (86%)	34 (187%)	53 (86%)	53 (86%)
Eyraud	1 (9%)	10 (121%)	9 (91%)	9 (91%)
Fager & Mc Gowan	6 (3%)	45 (42%)	6 (3%)	6 (3%)
Fleiss	9 (14%)	40 (95%)	9 (14%)	9 (14%)
Goodman & Kruskal	20 (5%)	44 (64%)	20 (5%)	20 (5%)
Hurlbert_1	33 (81%)	32 (147%)	33 (81%)	33 (81%)
Hurlbert_2	33 (81%)	32 (147%)	33 (81%)	33 (81%)
Johnson_1	9 (4%)	43 (83%)	9 (4%)	9 (4%)
Kuder & Richardson	33 (101%)	35 (510%)	33 (101%)	33 (101%)
Kulczynski_2	9 (4%)	43 (83%)	9 (4%)	9 (4%)
Legendre & Legendre	33 (11%)	39 (98%)	33 (11%)	33 (11%)
Loevinger	44 (73%)	37 (174%)	44 (73%)	44 (73%)
Mc Connaughey	9 (4%)	43 (83%)	9 (4%)	9 (4%)
Michael	44 (72%)	36 (159%)	44 (72%)	44 (72%)
Pearson & Heron	26 (197%)	26 (302%)	26 (197%)	26 (197%)
Pearson_2	32 (58%)	33 (98%)	32 (58%)	32 (58%)
Simpson	45 (66%)	34 (163%)	45 (66%)	45 (66%)
Sokal & Sneath_2	9 (1%)	47 (69%)	9 (1%)	9 (1%)
Sokal & Sneath_4	38 (20%)	36 (116%)	38 (20%)	38 (20%)
Tarwid	34 (76%)	34 (150%)	34 (76%)	34 (76%)
Tschuproff	30 (29%)	33 (94%)	30 (29%)	30 (29%)
Yule_2	42 (108%)	34 (218%)	42 (108%)	42 (108%)
Yule_3	41 (77%)	40 (159%)	41 (77%)	41 (77%)
Anónimo_5	20 (4%)	#	#	#
Gilbert & Wells	35 (41%)	#	#	#
Stiles	24 (60%)	#	#	#
Clement	*	*	*	*
Cole_3	*	*	*	*
Cole_4	*	*	*	*
Dice_1	*	*	*	*
Dice_2	*	*	*	*
Peirce_2	*	*	*	*
Peirce_3	*	*	*	*
Mountford	+	+	+	+
Johnson_2	@	@	@	@
Peirce_1	@	@	@	@
Kulczynski_1	@	@	@	@
Sokal & Sneath_3	@	@	@	@

Nota: el símbolo # indica que el índice es no acotado, el \* que es no simétrico, el + que es indeterminado y el @ una o más de las anteriores, en estos casos la propiedad no puede ser verificada. El sombreado corresponde a valores extremos de inercia.

## **CAPÍTULO 4**

### **SELECCIÓN DE LOS MEJORES ÍNDICES A PARTIR DE SUS PROPIEDADES**

Se estudian en esta tesis quince propiedades consideradas por varios autores (Hubálek 1982, Giansanti 1990, Warrens 2008) de interés en índices de similitud y disimilitud para datos binarios. Se presenta una colección de setenta y siete índices de similitud y tres disimilitudes derivadas en los casos posibles. La originalidad de este trabajo se debe a que se consideran todas las propiedades y todos los índices encontrados en la literatura. Por lo general los autores trabajan con un número reducido de ellos y estudian unas pocas propiedades.

Otros autores trabajan con conjuntos definidos de datos por lo que a veces hay discrepancias en los resultados. En este trabajo se utilizan matrices de datos de presencia/ausencia simuladas bajo distintos escenarios.

Se han encontrado en la literatura algunos errores en las fórmulas de los índices, lo que da lugar a discrepancias cuando se comparan los resultados.

Algunas propiedades que se estudian en esta tesis han sido enunciadas pero no probadas con datos reales o simulados.

Si bien existen muchos índices de similitud y disimilitud en la literatura, no se tiene certeza de cuáles son los que más propiedades cumplen y al momento de seleccionar uno de ellos no se sabe con exactitud cuál elegir. Es interesante el hecho de que esta discusión ya se daba por el 1900, cuando autores como Yule (1912) comparan la elección de un coeficiente de asociación con el uso de la media aritmética, geométrica y la mediana, y encuentran que todas las formas de promedio son medidas con propiedades análogas, pero no dan los mismos valores y que los distintos coeficientes que han sido sugeridos para medir asociación difieren precisamente en la misma forma.

Se presentan a continuación las conclusiones de las propiedades estudiadas para los setenta y siete índices de similitud y para cada una de las tres disimilitudes derivadas de los mismos.

En la Tabla 16 se presentan los resultados de las catorce propiedades estudiadas en los índices de similitud, no se presenta la Propiedad Métrica (Prop. 13) porque esta es exclusiva de las disimilitudes y para la Propiedad de Euclideanidad (Prop. 15) de las matrices de disimilitud, solamente se indica si la matriz de similitud es semi definida positiva. La Propiedad de Distribuciones en el muestreo (Prop.11) se presenta desdoblada en Prop.11a que corresponde al cumplimiento del supuesto de Normalidad, el cual no se cumple para ningún índice. Se agrega la información del cumplimiento de la asimetría o curtosis de una distribución normal. En la columna Prop.11b se agrega la información sobre la forma de la distribución empírica, específicamente si la misma se puede considerar una distribución Normal, teniendo en cuenta la información de la asimetría y curtosis en el sentido del gráfico de Cullen y Frey.

Tabla 16: Resumen de todas las propiedades estudiadas en los índices de Similitud (A)

Índice de Similitud	Intervalo	Prop.1	Prop.2	Prop.3a	Prop.3b	Prop.4	Prop.5	Prop.6	Prop.7(*)	Prop.8	Prop.9	Prop.10	Prop.11a	Prop.11b	Prop.12	Prop.14	Prop.15
Anónimo_1	[-1,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	curtosis	NO	SI	SI	SI
Anónimo_2	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	SI	SI	SI	curtosis	SI	SI	SI	SI
Anónimo_3	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	SI	SI	SI	curtosis	SI	SI	SI	SI
Anónimo_4	[0,1]	SI	SI	NO	SI	abs.	SI	SI	moderada	cóncavo	SI	SI	NO	NO	SI	NO	NO
Anónimo_5	[-1,+∞]	NO	SI	SI	NO	-----	NO	SI	moderada	SI	SI	SI	NO	NO	SI	NO	NO
Baroni-Urbani & Buser_1 (1976)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	cóncavo	SI	SI	NO	NO	NO	SI	NO
Baroni-Urbani & Buser_2 (1976)	[-1,1]	SI	SI	SI	SI	abs.	SI	SI	fuerte	cóncavo	SI	SI	NO	NO	NO	SI	NO
Braun-Blanquet (1932)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	asimetría	SI	NO	SI	SI
Clement (1976)	[0,n]	NO	SI	SI	SI	abs.	NO	NO	moderada	SI	SI	SI	NO	SI	SI	NO	-----
Clifford & Stephenson (1975)	[0,1]	SI	SI	NO	SI	abs.	NO	SI	fuerte	cóncavo	NO	NO	NO	NO	NO	NO	SI
Cohen (1960)	[-1,1]	SI	SI	NO/SI (b=c)	SI	abs.	NO	SI	fuerte	SI	SI	SI	curtosis	SI	SI	SI	SI
Cole_1 (1949)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	convexo	SI	SI	NO	NO	SI	NO	NO
Cole_2 (1949)	[0,1]	SI	SI	NO	SI	abs.	NO	SI	moderada	SI	NO	NO	NO	NO	NO	NO	NO
Cole_3 (1949)	[-∞,1]	NO	SI	NO/SI (b=c)	SI	compl.(b=0)/abs.(c=0)	NO	NO	moderada	SI	SI	SI	asimetría	NO	SI	NO	-----
Cole_4 (1949)	[-∞,1]	NO	SI	NO/SI (b=c)	SI	abs.(b=0)/compl.(c=0)	NO	NO	débil	SI	SI	SI	NO	NO	SI	NO	-----
Dennis (en Holliday et al. 2002)	[-∞,+∞]	NO	SI	NO	NO	-----	NO	SI	fuerte	SI	SI	SI	NO	SI	SI	NO	SI
Dice_1 (1945)	[0,1]	SI	SI	SI	SI	compl.(b=0)/abs.(c=0)	SI	NO	débil	SI	SI	SI	NO	NO	SI	NO	-----
Dice_2 (1945)	[0,1]	SI	SI	SI	SI	abs.(b=0)/compl.(c=0)	SI	NO	débil	SI	SI	SI	NO	NO	SI	NO	-----
Dice_3 (1945)	[0,2]	NO	SI	SI	NO/SI (a=1)	abs. (a=1)	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	SI
Digby (1983)	[-1,1]	SI	SI	SI	SI	compl.	SI	SI	moderada	convexo	SI	SI	NO	NO	SI	NO	NO
Doollittle (1885)	[0,1]	SI	SI	NO	SI	abs.	NO	SI	fuerte	cóncavo	NO	NO	NO	NO	NO	NO	SI
Eyraud (1936)	[-1,0]	NO/#	NO	NO/SI (bc=1)	NO	-----	NO	SI	NO	SI	SI	NO	NO	NO	SI	NO	NO
Fager & Mc Gowan (1963)	[-1,1]	SI/#	SI	NO	NO	-----	NO	SI	moderada	SI	SI	SI	NO	NO	SI	NO	NO
Faith (1983)	[0,1]	SI/#	SI	SI	NO	-----	NO	SI	moderada	SI	SI	SI	curtosis	NO	SI	SI	SI
Fleiss (1975)	[-∞,1]	NO	SI	NO	SI	abs.	NO	SI	fuerte	SI	SI	SI	NO	NO	SI	NO	NO
Forbes_1 (1907)	[0,+∞]	NO	SI	SI	NO	-----	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	SI
Forbes_2 (1907)	[-1,1]	SI/#	SI	NO/SI (b=c)	SI	abs.	NO	SI	fuerte	SI	SI	SI	NO	NO	NO	SI	SI
Forbes_3 (en Cole 1949)	[-1,+∞]	NO	SI	SI	NO	-----	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	SI
Fossium (en Holliday et al. 2002)	[0,+∞]	NO	SI	NO	NO	-----	NO	SI	moderada	convexo	SI	SI	curtosis	NO	NO	NO	SI
Gilbert & Wells (1966)	[-∞,+∞]	NO	SI	SI	NO	-----	SI	SI	moderada	cóncavo	-----	SI	asimetría	NO	NO	NO	NO
Goodall (1967)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	cóncavo	SI	SI	NO	NO	SI	SI	SI
Goodman & Kruskal (1954)	[-1,1]	SI	SI	SI	SI	abs.	SI	SI	fuerte	SI	SI	SI	curtosis	SI	NO	SI	NO
Hamann (1961)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	asimetría	SI	SI	SI	SI
Harris & Lahey (1978)	[0,n]	NO	SI	SI	SI	abs.	NO	SI	moderada	convexo	SI	SI	NO	NO	SI	SI	SI
Hawkins & Dotson (1968)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	convexo	SI	SI	NO	NO	SI	SI	SI
Hurlbert_1 (1969)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	curtosis	SI	NO	NO	NO
Hurlbert_2 (1969)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	curtosis	SI	NO	NO	NO
Jaccard (1901)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	convexo	SI	SI	curtosis	NO	SI	SI	SI
Johnson_1 (1964)	[0,2]	NO	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	NO
Johnson_2 (1971)	[0,+∞]	NO	-----	SI	SI	compl.(b=0)/abs.(c=0)	SI	NO	débil	convexo	NO	SI	NO	NO	SI	NO	-----
Kendall (en Podani 1997)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	cóncavo	SI	SI	NO	NO	SI	SI	SI
Kuder & Richardson (1937)	[-∞,1]	NO	SI	SI	SI	abs.	NO	SI	débil	cóncavo	NO	SI	NO	NO	SI	NO	NO
Kulczynski_1 (1927)	[0,+∞]	NO	-----	SI	SI	abs.	SI	SI	moderada	convexo	NO	SI	NO	NO	SI	NO	-----
Kulczynski_2 (1927)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	NO
Lamont & Grant (1979)	[0,0.5]	NO	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	curtosis	NO	SI	SI	SI
Legendre & Legendre (1998)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	débil	cóncavo	SI	SI	NO	NO	SI	SI	NO
Loewinger (1948)	[-∞,1]	NO	SI	NO	SI	compl.	NO	SI	fuerte	SI	SI	SI	NO	NO	NO	NO	NO
Maxwell & Pillmer (1968)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	NO	SI	SI	SI	SI
Mc Connaughey (1964)	[-1,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	NO	NO	SI	NO	NO
Michael (1920)	[-1,1]	SI	NO	NO/SI (b=c)	NO/SI (a=d)	abs. (a=d)	NO	SI	fuerte	convexo	SI	SI	NO	NO	NO	SI	NO
Mountford (1962)	[0,2]	NO	-----	SI	NO	-----	SI	SI	moderada	convexo	NO	SI	NO	NO	NO	NO	-----
Ochiai (1957)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	curtosis	NO	SI	NO	SI
Pearson & Heron (1913)	[-1,1]	SI	SI	SI	SI	compl.	SI	SI	débil	cónc./conv.	NO	NO	NO	NO	NO	NO	NO
Pearson_1 (1905)	[0,n]	NO	SI	NO	SI	abs.	NO	SI	fuerte	cóncavo	NO	NO	NO	NO	NO	NO	SI
Pearson_2 (1905)	[0,1]	SI/#	SI	NO	NO	-----	NO	SI	fuerte	cóncavo	NO	NO	NO	NO	NO	NO	NO
Peirce_1 (1884)	[0,1]	SI	-----	NO	NO	-----	NO	NO	NO	SI	SI	NO	NO	NO	SI	NO	-----
Peirce_2 (1884)	[-1,1]	SI	SI	SI	SI	abs.	NO	NO	fuerte	SI	SI	SI	NO	SI	SI	NO	-----
Peirce_3 (1884)	[-1,1]	SI	SI	SI	SI	abs.	NO	NO	moderada	SI	SI	SI	NO	SI	SI	NO	-----
Rogers & Tanimoto (1960)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	convexo	SI	SI	NO	NO	SI	SI	SI
Rogot & Goldberg (1966)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	SI	SI	SI	NO	SI	SI	SI	SI
Russell & Rao (1940)	[0,1]	SI/#	SI	SI	NO	-----	SI	SI	débil	SI	SI	SI	curtosis	NO	SI	SI	SI
Scott (1955)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	NO	SI	SI	SI	SI
Simpson (1943)	[0,1]	SI	SI	SI	SI	compl.	SI	SI	débil	SI	SI	SI	curtosis	NO	NO	NO	NO
Sokal & Michener (1958)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	SI	SI	SI	asimetría	SI	SI	SI	SI
Sokal & Sneath_1 (1963)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	convexo	SI	SI	NO	NO	SI	SI	SI
Sokal & Sneath_2 (1963)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	SI	SI	SI	NO	SI	SI	NO	NO
Sokal & Sneath_3 (1963)	[0,+∞]	NO	-----	SI	SI	abs.	NO	SI	moderada	convexo	NO	SI	NO	NO	SI	NO	-----
Sokal & Sneath_4 (1963)	[0,1]	SI	SI	SI	SI	abs.	NO	SI	moderada	cóncavo	SI	SI	NO	NO	SI	SI	NO
Sokal & Sneath_5 (1963)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	convexo	SI	SI	curtosis	NO	SI	NO	SI
Sorensen (1948)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	SI	SI	SI	curtosis	NO	SI	SI	SI
Sorgenfrei (1959)	[0,1]	SI	SI	SI	SI	abs.	SI	SI	moderada	convexo	SI	SI	NO	NO	NO	NO	SI
Stiles (1946)	[-∞,+∞]	NO	SI	NO	NO	-----	NO	SI	débil	cóncavo	NO	NO	NO	NO	NO	NO	NO
Tarwid (1960)	[-1,1]	SI/#	SI	SI	NO	-----	SI	SI	moderada	cóncavo	SI	SI	NO	NO	NO	NO	NO
Tschuproff (en Rodriguez-Salazar 2001)	[0,1]	SI	SI	NO	SI	abs.	NO	SI	fuerte	SI	NO	NO	NO	NO	NO	NO	NO
Yule_1 (1912)	[-1,1]	SI	SI	SI	SI	abs.	NO	SI	fuerte	SI	SI	SI	NO	SI	SI	NO	SI
Yule_2 (1900)	[-1,1]	SI	SI	SI	SI	compl.	SI	SI	fuerte	convexo	SI	SI	NO	NO	SI	NO	NO
Yule_3 (1912)	[-1,1]	SI	SI	SI	SI	compl.	SI	SI	fuerte	cóncavo	SI	SI	NO	NO	NO	NO	NO

Nota: en la Prop.1 el símbolo # significa que el índice no alcanza los límites. En la Prop. 7, fuerte corresponde a una correlación r en el intervalo [0.90; 1], moderada en [0.50; 0.90) y para débil r<0.50.

Se seleccionan los mejores índices, es decir, aquellos que cumplen todas o casi todas las propiedades definidas por Hubálek (1982) y estudiadas en este trabajo. Tal como lo definiera Hubálek (1982) en su trabajo, hay cinco propiedades (2, 3, 6, 7 y 10) que son las condiciones obligatorias que deben cumplir los índices para ser considerados “buenos índices“. Las restantes son opcionales, sin embargo alguna de las dos propiedades distribucionales y que la matriz de similitud sea psd son muy importantes. Ningún índice cumple con el 100% de las propiedades, se detalla a continuación los que fallan solo en una, en dos o en tres de las propiedades opcionales.

#### Índices que no cumplen con solo una propiedad:

Existen dos índices que no cumplen con la segunda de las propiedades distribucionales (11b), ellos son **Anónimo\_1** y **Sorensen**. El primero varía entre  $[-1,1]$  y el segundo entre  $[0,1]$ , ambos consideran en su fórmula solamente las cantidades  $a$ ,  $b$  y  $c$  y no tienen en cuenta las ausencias conjuntas. Cabe mencionar que el primero de ellos, si bien figura como anónimo en la tesis de Warrens (2008), se ha encontrado una referencia del índice a Van der Maarel (1969) en el paquete Simba de R (Jurasinski 2012).

Los índices de **Anónimo\_2**, **Anónimo\_3**, **Hamann** y **Sokal & Michener** no cumplen con la Propiedad 5, que establece que un índice debe alcanzar también el valor mínimo cuando  $a=0$ , y no solamente cuando  $a=d=0$  (propiedad 3a). Esta propiedad es un poco más fuerte que la propiedad 3a). **Anónimo\_2** varía entre  $[0,1]$ , **Anónimo\_3** y **Hamann** varían entre  $[-1,1]$  y **Sokal & Michener** entre  $[0,1]$ . Los dos primeros índices contienen en su fórmula las cantidades  $b$ ,  $c$  y  $d$ , por lo tanto no serían tan recomendables para datos de vegetación, mientras que los dos últimos tienen en cuenta las cuatro cantidades  $a$ ,  $b$ ,  $c$  y  $d$ . Cabe aclarar que los cuatro índices, si bien no cumplen con la primera propiedad de normalidad, los dos primeros tienen kurtosis similar a la normal, mientras que los dos últimos tienen asimetría similar a la distribución normal.

El índice de **Braun-Blanquet** no cumple con la propiedad de insesgamiento, pero en la Tabla 12, se observa que el porcentaje de valores insesgados encontrados en los ensayos realizados es alto (85%). Este índice cumple

además con la asimetría de una distribución Normal, varía entre  $[0,1]$  y solo considera las cantidades  $a$ ,  $b$  y  $c$  en su fórmula.

#### Índices que no cumplen con dos propiedades:

El índice de **Jaccard** no cumple con la segunda propiedad de distribución (11b), pero tiene kurtosis similar a una distribución normal (Propiedad 11a) y no cumple con la propiedad de linealidad ya que es convexo. Tiene el menor valor de CV de los índices convexos (Tabla 10), siendo uno de los más recomendables (Gower & Legendre 1986). Este índice varía entre  $[0,1]$  y no considera las ausencias conjuntas en su fórmula y es uno de los más utilizados en Ecología de comunidades, junto con el índice de Sorensen.

El índice de **Ochiai** tampoco cumple con la Propiedad 11b de distribución, pero sí con la 11a, ya que tiene kurtosis similar a una distribución Normal y en ciertas ocasiones se hace indeterminado. Sin embargo, esto se puede subsanar redefiniendo al índice como recomienda Warrens (2008). Varía entre  $[0,1]$  y no considera en su fórmula las ausencias conjuntas, siendo por lo tanto apropiado para datos de Ecología.

Dos índices, **Rogot & Goldberg** y **Scott**, no cumplen con la propiedad distribucional (11a), sin embargo cumplen con 11b y tampoco alcanzan el mínimo cuando  $a=0$  (Propiedad 5), que resulta más estricta que la Propiedad 3a. El primero varía entre  $[0,1]$  y el segundo entre  $[-1,1]$ . Ambos consideran en su fórmula las cuatro cantidades  $a$ ,  $b$ ,  $c$  y  $d$ .

El índice de **Lamont & Grant** no cumple con los límites considerados aceptables por Hubálek (1982), el mismo varía entre  $[0, 0.5]$  y no cumple la propiedad distribucional 11a pero sí la 11b. No considera las ausencias conjuntas en su fórmula por lo que sería apropiado para datos de vegetación.

El índice de **Maxwell & Pilliner** por su parte no cumple la propiedad distribucional 11a pero sí la 11b. Tampoco cumple con la propiedad 5. Varía entre  $[-1,1]$  y contiene en su fórmula las cuatro cantidades  $a$ ,  $b$ ,  $c$  y  $d$ .

Existe un grupo de índices que no cumplen con dos propiedades, pero algunas de ellas son obligatorias o se trata de las dos propiedades distribucionales o la euclideanidad, ellos son: Cohen (no cumple con la prop.3a, alcanzar el valor mínimo cuando  $a=d=0$  y con la prop.5, alcanzar el mínimo pero solamente si

$a=0$ ). Por su parte Faith no cumple con las prop.3b, alcanzar el valor máximo cuando  $b=c=0$  y con la prop.5. Goodman & Kruskal no cumple con las propiedades de insesgamiento y euclideanidad, siendo esta última importante, Russell & Rao no cumple con la prop.3b, y tampoco con la propiedad distribucional 11b, pero si la 11a y Sokal & Sneath\_1 no cumple ninguna de las dos propiedades distribucionales.

#### Índices que no cumplen con tres propiedades:

El índice de **Yule\_1**, no cumple la propiedad 5, la primera propiedad de distribución (11a) y a veces se hace indeterminado. Varía entre  $[-1,1]$  y considera en la fórmula los valores  $a, b, c$  y  $d$ , se le conoce como el coeficiente Phi ( $\Phi$ ).

Por su parte el índice de **Sokal & Sneath\_5** tampoco cumple la primera propiedad de distribución (11a), a veces se hace indeterminado y además es convexo, pero al igual que el índice de Jaccard, si observamos la Tabla 10, vemos que tiene el segundo menor valor de CV de los índices convexos, lo cual lo hace recomendable. Varía entre  $[0,1]$  y considera en la fórmula los valores  $a, b, c$  y  $d$ .

El resto de los índices no cumplen con tres o más de las propiedades, siendo alguna o algunas de ellas obligatorias. El peor desempeño se encontró para el índice de Stiles, que no cumple con doce de las catorce propiedades estudiadas para la similitud.

Tabla 17: Resumen de todas las propiedades estudiadas en los índices de Disimilitud 1 (D1)

Índice de similitud	Intervalo	Prop.2	Prop.6	Prop.11a	Prop.11b	Prop.13	Prop.14	Prop.15
Anónimo_1	[0,1]	SI	SI	curtosis	NO	NO	SI	NO
Anónimo_2	[0,1]	SI	SI	curtosis	SI	NO	SI	NO
Anónimo_3	[0,1]	SI	SI	curtosis	SI	NO	SI	NO
Anónimo_4	[0,1]	SI	SI	NO	NO	NO	NO	NO
Anónimo_5	--	-	-	-	-	-	-	-
Baroni-Urbani & Buser_1 (1976)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Baroni-Urbani & Buser_2 (1976)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Braun-Blanquet (1932)	[0,1]	SI	SI	asimetría	SI	SI	SI	NO
Clement (1976)	[0,n]	SI	NO	NO	SI	NO	NO	-
Clifford & Stephenson (1975)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Cohen (1960)	[0,1]	SI	SI	curtosis	SI	NO	SI	NO
Cole_1 (1949)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Cole_2 (1949)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Cole_3 (1949)	[0,+∞]	SI	NO	asimetría	NO	-	NO	-
Cole_4 (1949)	[0,+∞]	SI	NO	NO	NO	-	NO	-
Dennis (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Dice_1 (1945)	[0,1]	SI	NO	NO	NO	-	NO	-
Dice_2 (1945)	[0,1]	SI	NO	NO	NO	-	NO	-
Dice_3 (1945)	[0,2]	SI	SI	NO	NO	SI	NO	NO
Digby (1983)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Doolittle (1885)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Eyraud (1936)	[0,1]	NO	SI	NO	NO	SI	NO	NO
Fager & Mc Gowan (1963)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Faith (1983)	[0,1]	SI	SI	curtosis	NO	SI	SI	NO
Fleiss (1975)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Forbes_1 (1907)	--	-	-	-	-	-	-	-
Forbes_2 (1907)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Forbes_3 (en Cole 1949)	--	-	-	-	-	-	-	-
Fossum (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Gilbert & Wells (1966)	--	-	-	-	-	-	-	-
Goodall (1967)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Goodman & Kruskal (1954)	[0,1]	SI	SI	curtosis	SI	NO	SI	NO
Hamann (1961)	[0,1]	SI	SI	asimetría	SI	SI	SI	NO
Harris & Lahey (1978)	[0,n]	SI	SI	NO	NO	NO	SI	NO
Hawkins & Dotson (1968)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Hurlbert_1 (1969)	[0,1]	SI	SI	curtosis	SI	NO	NO	NO
Hurlbert_2 (1969)	[0,1]	SI	SI	curtosis	SI	NO	NO	NO
Jaccard (1901)	[0,1]	SI	SI	curtosis	NO	SI	SI	NO
Johnson_1 (1964)	[0,2]	SI	SI	NO	NO	NO	NO	NO
Johnson_2 (1971)	--	-	-	-	-	-	-	-
Kendall (1997)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Kuder & Richardson (1937)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Kulczynski_1 (1927)	--	-	-	-	-	-	-	-
Kulczynski_2 (1927)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Lamont & Grant (1979)	[0,0.5]	SI	SI	curtosis	NO	NO	SI	NO
Legendre & Legendre (1998)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Loevinger (1948)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Maxwell & Pilliner (1968)	[0,1]	SI	SI	NO	SI	NO	SI	NO
Mc Connaughey (1964)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Michael (1920)	[0,1]	NO	SI	NO	NO	NO	SI	NO
Mountford (1962)	[0,2]	-	SI	NO	NO	-	NO	-
Ochiai (1957)	[0,1]	SI	SI	curtosis	NO	NO	NO	NO
Pearson & Heron (1913)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Pearson_1 (1905)	[0,n]	SI	SI	NO	NO	NO	NO	NO
Pearson_2 (1905)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Peirce_1 (1884)	[0,1]	-	NO	NO	NO	-	NO	-
Peirce_2 (1884)	[0,1]	SI	NO	NO	SI	-	NO	-
Peirce_3 (1884)	[0,1]	SI	NO	NO	SI	-	NO	-
Rogers & Tanimoto (1960)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Rogot & Goldberg (1966)	[0,1]	SI	SI	NO	SI	NO	SI	NO
Russell & Rao (1940)	[0,1]	SI	SI	curtosis	NO	SI	SI	NO
Scott (1955)	[0,1]	SI	SI	NO	SI	NO	SI	NO
Simpson (1943)	[0,1]	SI	SI	curtosis	NO	NO	NO	NO
Sokal & Michener (1958)	[0,1]	SI	SI	asimetría	SI	SI	SI	NO
Sokal & Sneath_1 (1963)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Sokal & Sneath_2 (1963)	[0,1]	SI	SI	NO	SI	NO	NO	NO
Sokal & Sneath_3 (1963)	--	-	-	-	-	-	-	-
Sokal & Sneath_4 (1963)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Sokal & Sneath_5 (1963)	[0,1]	SI	SI	curtosis	NO	NO	NO	NO
Sorensen (1948)	[0,1]	SI	SI	curtosis	NO	NO	SI	NO
Sorgenfrei (1959)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Stiles (1946)	--	-	-	-	-	-	-	-
Tarwid (1960)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Tschuproff (en Rodriguez-Salazar 2001)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Yule_1 (1912)	[0,1]	SI	SI	NO	SI	NO	NO	NO
Yule_2 (1900)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Yule_3 (1912)	[0,1]	SI	SI	NO	NO	NO	NO	NO

Con respecto a la Disimilitud 1, se concluyó que ninguno de los índices cumple la propiedad de euclideanidad (prop. 15), por lo tanto, no se recomienda utilizar la Disimilitud 1 ( $Dis.1 = 1 - A$ ) para representar puntos en un espacio de dimensión reducida.

Los índices de disimilitud que cumplen todas las propiedades excepto la euclideanidad son los siguientes: **Braun-Blanquet**, **Hamann** y **Sokal & Michener**. Los índices de **Faith Jaccard**, y **Russell & Rao** no cumplen la propiedad distribucional 11b pero si la 11a. Otra propiedad que cumplen muy pocos índices de Disimilitud 1 es la métrica, solamente el 18%. Los índices de **Anónimo\_2**, **Anónimo\_3**, **Cohen y Goodman & Kruskal** no la cumplen y por lo tanto no hacen posible un ordenamiento adecuado de puntos en un espacio métrico. Los índices de **Anónimo\_1**, **Lamont & Grant**, **Maxwell & Pilliner**, **Rogot & Goldberg**, **Scott** y **Sorensen** no cumplen una de las dos propiedades distribucionales y tampoco la propiedad métrica.

Las demás disimilitudes por no cumplir tres o más propiedades y entre ellas las dos propiedades distribucionales, o algunas de las propiedades obligatorias, como ser la simetría no son seleccionadas como medidas recomendables.

Tabla 18: Resumen de todas las propiedades estudiadas en los índices de Disimilitud 2 (D2)

Índice de similitud	Intervalo	Prop.2	Prop.6	Prop.11a	Prop.11b	Prop.13	Prop.14	Prop.15
Anónimo_1	[0,1]	SI	SI	NO	NO	SI	SI	SI
Anónimo_2	[0,1]	SI	SI	NO	NO	SI	SI	SI
Anónimo_3	[0,1]	SI	SI	NO	NO	SI	SI	SI
Anónimo_4	[0,1]	SI	SI	NO	NO	NO	NO	NO
Anónimo_5	--	-	-	-	-	-	-	-
Baroni-Urbani & Buser_1 (1976)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Baroni-Urbani & Buser_2 (1976)	[0,1]	SI	SI	NO	NO	NO	SI	NO
Braun-Blanquet (1932)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Clement (1976)	[0,vn]	SI	NO	NO	NO	-	NO	-
Clifford & Stephenson (1975)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Cohen (1960)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Cole_1 (1949)	[0,1]	SI	SI	curtosis	NO	NO	NO	NO
Cole_2 (1949)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Cole_3 (1949)	[0,+∞]	SI	NO	NO	NO	-	NO	-
Cole_4 (1949)	[0,+∞]	SI	NO	NO	NO	-	NO	-
Dennis (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Dice_1 (1945)	[0,1]	SI	NO	NO	NO	-	NO	-
Dice_2 (1945)	[0,1]	SI	NO	curtosis	NO	-	NO	-
Dice_3 (1945)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Digby (1983)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Doolittle (1885)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Eyraud (1936)	[0,1]	NO	SI	NO	NO	SI	NO	NO
Fager & Mc Gowan (1963)	[0,1]	SI	SI	NO	SI	SI	NO	NO
Faith (1983)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Fleiss (1975)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Forbes_1 (1907)	--	-	-	-	-	-	-	-
Forbes_2 (1907)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Forbes_3 (en Cole 1949)	--	-	-	-	-	-	-	-
Fossum (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Gilbert & Wells (1966)	--	-	-	-	-	-	-	-
Goodall (1967)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Goodman & Kruskal (1954)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Hamann (1961)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Harris & Lahey (1978)	[0,vn]	SI	SI	NO	NO	SI	SI	SI
Hawkins & Dotson (1968)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Hurlbert_1 (1969)	[0,1]	SI	SI	curtosis	NO	NO	NO	NO
Hurlbert_2 (1969)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Jaccard (1901)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Johnson_1 (1964)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Johnson_2 (1971)	--	-	-	-	-	-	-	-
Kendall (1997)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Kuder & Richardson (1937)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Kulczynski_1 (1927)	--	-	-	-	-	-	-	-
Kulczynski_2 (1927)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Lamont & Grant (1979)	[0,v0.5]	SI	SI	NO	NO	SI	SI	SI
Legendre & Legendre (1998)	[0,1]	SI	SI	NO	SI	SI	SI	NO
Loevinger (1948)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Maxwell & Pilliner (1968)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Mc Connaughey (1964)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Michael (1920)	[0,1]	NO	SI	NO	NO	NO	SI	NO
Mountford (1962)	[0,v2]	-	SI	NO	NO	NO	NO	-
Ochiai (1957)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Pearson & Heron (1913)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Pearson_1 (1905)	[0,vn]	SI	SI	NO	NO	SI	NO	SI
Pearson_2 (1905)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Peirce_1 (1884)	[0,1]	-	NO	NO	NO	-	NO	-
Peirce_2 (1884)	[0,1]	SI	NO	NO	NO	-	NO	-
Peirce_3 (1884)	[0,1]	SI	NO	NO	NO	-	NO	-
Rogers & Tanimoto (1960)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Rogot & Goldberg (1966)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Russell & Rao (1940)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Scott (1955)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Simpson (1943)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Sokal & Michener (1958)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Sokal & Sneath_1 (1963)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Sokal & Sneath_2 (1963)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Sokal & Sneath_3 (1963)	--	-	-	-	-	-	-	-
Sokal & Sneath_4 (1963)	[0,1]	SI	SI	NO	NO	SI	SI	NO
Sokal & Sneath_5 (1963)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Sorensen (1948)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Sorgenfrei (1959)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Stiles (1946)	--	-	-	-	-	-	-	-
Tarwid (1960)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Tschuproff (en Rodriguez-Salazar 2001)	[0,1]	SI	SI	NO	NO	SI	NO	NO
Yule_1 (1912)	[0,1]	SI	SI	NO	NO	SI	NO	SI
Yule_2 (1900)	[0,1]	SI	SI	NO	NO	NO	NO	NO
Yule_3 (1912)	[0,1]	SI	SI	NO	NO	NO	NO	NO

Con respecto a la Disimilitud 2, la mayoría de los índices no cumplen con las dos propiedades distribucionales, 96% y 97% respectivamente. Se encuentran en este trabajo veintidós índices que cumplen con el resto de las propiedades. Ellos son: **Anónimo\_1, Anónimo\_2, Anónimo\_3, Braun-Blanquet, Cohen, Faith, Forbes\_2, Goodall, Hamann, Harris & Lahey, Hawkins & Dotson, Jaccard, Kendall, Lamont & Grant, Maxwell & Pilliner, Roger & Tanimoto, Rogot & Goldberg, Russell & Rao, Scott, Sokal & Michener, Sokal & Sneath\_1 y Sorensen.**

Algunos pocos índices (10%) además de no cumplir las propiedades distribucionales, también se hacen indeterminados para algunos valores de  $a$ ,  $b$ ,  $c$  y/o  $d$ : **Clifford & Stephenson, Dice\_3, Doolittle, Ochiai, Pearson\_1, Sokal & Sneath\_5, Sorgenfrei y Yule\_1.**

El resto de los índices no cumple con dos o más propiedades, pero las mismas se refieren a la simetría que consideramos obligatoria o la métrica o euclideanidad, que resultan propiedades importantes para las medidas de disimilitud, por lo tanto no son seleccionadas en este trabajo.

Tabla 19: Resumen de todas las propiedades estudiadas en los índices de Disimilitud 3 (D3)

Índice de similitud	Intervalo	Prop.2	Prop.6	Prop.11a	Prop.11b	Prop.13	Prop.14	Prop.15
Anónimo_1	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Anónimo_2	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Anónimo_3	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Anónimo_4	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Anónimo_5	--	-	-	-	-	-	-	-
Baroni-Urbani & Buser_1 (1976)	[0,v2]	SI	SI	NO	NO	NO	SI	NO
Baroni-Urbani & Buser_2 (1976)	[0,v2]	SI	SI	NO	NO	NO	SI	NO
Braun-Blanquet (1932)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Clement (1976)	[0,v(2n)]	SI	NO	NO	NO	-	NO	-
Clifford & Stephenson (1975)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Cohen (1960)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Cole_1 (1949)	[0,v2]	SI	SI	curtosis	NO	NO	NO	NO
Cole_2 (1949)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Cole_3 (1949)	[0,+∞]	SI	NO	NO	NO	-	NO	-
Cole_4 (1949)	[0,+∞]	SI	NO	NO	NO	-	NO	-
Dennis (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Dice_1 (1945)	[0,v2]	SI	NO	NO	NO	-	NO	-
Dice_2 (1945)	[0,v2]	SI	NO	curtosis	NO	-	NO	-
Dice_3 (1945)	[0,2]	SI	SI	NO	NO	SI	NO	SI
Digby (1983)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Doolittle (1885)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Eyraud (1936)	[0,v2]	NO	SI	NO	NO	SI	NO	NO
Fager & Mc Gowan (1963)	[0,v2]	SI	SI	NO	SI	SI	NO	NO
Faith (1983)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Feiss (1975)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Forbes_1 (1907)	--	-	-	-	-	-	-	-
Forbes_2 (1907)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Forbes_3 (en Cole 1949)	--	-	-	-	-	-	-	-
Fossum (en Holliday et al. 2002)	--	-	-	-	-	-	-	-
Gilbert & Wells (1966)	--	-	-	-	-	-	-	-
Goodall (1967)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Goodman & Kruskal (1954)	[0,v2]	SI	SI	NO	NO	SI	SI	NO
Hamann (1961)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Harris & Lahey (1978)	[0,v(2n)]	SI	SI	NO	NO	SI	SI	SI
Hawkins & Dotson (1968)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Hurlbert_1 (1969)	[0,v2]	SI	SI	curtosis	NO	NO	NO	NO
Hurlbert_2 (1969)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Jaccard (1901)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Johnson_1 (1964)	[0,2]	SI	SI	NO	NO	NO	NO	NO
Johnson_2 (1971)	--	-	-	-	-	-	-	-
Kendall (1997)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Kuder & Richardson (1937)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Kulczynski_1 (1927)	--	-	-	-	-	-	-	-
Kulczynski_2 (1927)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Lamont & Grant (1979)	[0,1]	SI	SI	NO	NO	SI	SI	SI
Legendre & Legendre (1998)	[0,v2]	SI	SI	NO	SI	SI	SI	NO
Loevinger (1948)	[0,+∞]	SI	SI	NO	NO	NO	NO	NO
Maxwell & Pilliner (1968)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Mc Connaughey (1964)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Michael (1920)	[0,v2]	NO	SI	NO	NO	NO	SI	NO
Mountford (1962)	[0,2]	-	SI	NO	NO	NO	NO	-
Ochiai (1957)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Pearson & Heron (1913)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Pearson_1 (1905)	[0,v(2n)]	SI	SI	NO	NO	SI	NO	SI
Pearson_2 (1905)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Peirce_1 (1884)	[0,v2]	-	NO	NO	NO	-	NO	-
Peirce_2 (1884)	[0,v2]	SI	NO	NO	NO	-	NO	-
Peirce_3 (1884)	[0,v2]	SI	NO	NO	NO	-	NO	-
Rogers & Tanimoto (1960)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Rogot & Goldberg (1966)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Russell & Rao (1940)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Scott (1955)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Simpson (1943)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Sokal & Michener (1958)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Sokal & Sneath_1 (1963)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Sokal & Sneath_2 (1963)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Sokal & Sneath_3 (1963)	--	-	-	-	-	-	-	-
Sokal & Sneath_4 (1963)	[0,v2]	SI	SI	NO	NO	SI	SI	NO
Sokal & Sneath_5 (1963)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Sorensen (1948)	[0,v2]	SI	SI	NO	NO	SI	SI	SI
Sorgenfrei (1959)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Stiles (1946)	--	-	-	-	-	-	-	-
Tarwid (1960)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Tschuproff (en Rodriguez-Salazar 2001)	[0,v2]	SI	SI	NO	NO	SI	NO	NO
Yule_1 (1912)	[0,v2]	SI	SI	NO	NO	SI	NO	SI
Yule_2 (1900)	[0,v2]	SI	SI	NO	NO	NO	NO	NO
Yule_3 (1912)	[0,v2]	SI	SI	NO	NO	NO	NO	NO

Se presenta la Tabla 19, sin embargo la conclusión es la misma que para la Tabla 18, el único cambio son los límites de algunos de los índices de disimilitud.

Para unificar y sintetizar los resultados para similitudes y sus disimilitudes derivadas se presenta la Tabla 20 con un listado ordenado de todos los índices seleccionados, según el número de propiedades que satisfacen. Para la similitud A se estudian catorce propiedades mientras que para las disimilitudes D1, D2 y D3 se estudian seis.

*Tabla 20: selección de los mejores índices de Similitud (A) y Disimilitudes (D1, D2 y D3)*

Autores	Número de propiedades que cumplen los índices			
	Similitud (A)	Disimilitud (D1)	Disimilitud (D2)	Disimilitud (D3)
<b>Anónimo_1</b>	<b>13</b>	<b>3</b>	<b>5</b>	<b>5</b>
<b>Anónimo_2</b>	<b>13</b>	<b>4</b>	<b>5</b>	<b>5</b>
<b>Anónimo_3</b>	<b>13</b>	<b>4</b>	<b>5</b>	<b>5</b>
<b>Braun-Blanquet</b>	<b>13</b>	<b>5</b>	<b>5</b>	<b>5</b>
Clifford & Stephenson			4	4
Cohen		4	5	5
Dice_3			4	4
Doolittle			4	4
Faith		4	5	5
Forbes_2			5	5
Goodall			5	5
Goodman & Kruskal		4		
<b>Hamann</b>	<b>13</b>	<b>5</b>	<b>5</b>	<b>5</b>
Harris & Lahey			5	5
Hawkins & Dotson			5	5
<b>Jaccard</b>	<b>12</b>	<b>4</b>	<b>5</b>	<b>5</b>
Kendall			5	5
<b>Lamont &amp; Grant</b>	<b>12</b>	<b>3</b>	<b>5</b>	<b>5</b>
<b>Maxwell &amp; Pilliner</b>	<b>12</b>	<b>3</b>	<b>5</b>	<b>5</b>
Ochiai	12		4	4
Pearson_1			4	4
Roger & Tanimoto			5	5
<b>Rogot &amp; Goldberg</b>	<b>12</b>	<b>3</b>	<b>5</b>	<b>5</b>
Russell & Rao		4	5	5
<b>Scott</b>	<b>12</b>	<b>3</b>	<b>5</b>	<b>5</b>
<b>Sokal &amp; Michener</b>	<b>13</b>	<b>5</b>	<b>5</b>	<b>5</b>
Sokal & Sneath_1			5	5
Sokal & Sneath_5	11		4	4
<b>Sorensen</b>	<b>13</b>	<b>3</b>	<b>5</b>	<b>5</b>
Sorgenfrei			4	4
Yule_1	11		4	4

De esta forma, se eligen los mejores índices de similitud y disimilitudes derivadas que cumplen doce o más propiedades para la similitud y tres o más propiedades para las disimilitudes y que fueron seleccionados en las Tablas 16, 17, 18 y 19. Los mismos se encuentran remarcados en negritas y resaltados con un sombreado en la Tabla 20. Se concluye que los mejores índices encontrados en este trabajo son:

<p><b>Anónimo_1, se lo atribuyen a Van der Maarel (1969) en el paquete Simba del programa R</b></p> $A_1 = \frac{2a - (b + c)}{2d + (b + c)}$	<p>Rango de variación <math>[-1, 1]</math> No cumple una de las dos propiedades distribucionales</p>
<p><b>Anónimo_2</b></p> $A_2 = \frac{2d}{2d + b + c}$	<p>Rango de variación <math>[0, 1]</math> No cumple la propiedad de alcanzar el mínimo cuando <math>a=0</math></p>
<p><b>Anónimo_3</b></p> $A_3 = \frac{2d - (b + c)}{2d + (b + c)}$	<p>Rango de variación <math>[-1, 1]</math> No cumple la propiedad de alcanzar el mínimo cuando <math>a=0</math></p>
<p><b>Braun-Blanquet (1932)</b></p> $A_8 = \frac{a}{\text{máx.}\{(a + b), (a + c)\}}$	<p>Rango de variación <math>[0, 1]</math> No cumple la propiedad de ser insesgado</p>
<p><b>Hamann (1961); en psicología es conocido como índice G (Holley &amp; Guilford 1964), Hubert (1977)</b></p> $A_{33} = \frac{a + d - b - c}{n} = \frac{a + d - b - c}{a + d + b + c}$	<p>Rango de variación <math>[-1, 1]</math> No cumple la propiedad de alcanzar el mínimo cuando <math>a=0</math></p>

<p><b>Jaccard (1901), Sneath (1957), Iverson (1954), Agrell (1945), coeficiente de comunidad</b></p> $A_{38} = \frac{a}{a + b + c}$	<p>Rango de variación [0,1] No cumple una de las dos propiedades distribucionales y no es lineal (convexo)</p>
---	--

<p><b>Lamont &amp; Grant (1979)</b></p> $A_{45} = \frac{a}{2a + b + c}$	<p>Rango de variación [0,0.5] No cumple una de las dos propiedades distribucionales y no tiene límites convencionales</p>
---	---

<p><b>Maxwell &amp; Pilliner (1968)</b></p> $A_{48} = \frac{2(ad - bc)}{(a + b)(c + d) + (a + c)(b + d)}$	<p>Rango de variación [-1,1] No cumple una de las dos propiedades distribucionales y no alcanza el mínimo cuando a=0</p>
---	--

<p><b>Rogot &amp; Goldberg (1966) (coincide con la media aritmética de Anónimo_1 y Sorensen)</b></p> $A_{60} = \frac{a}{2a + b + c} + \frac{d}{2d + b + c}$	<p>Rango de variación [0,1] No cumple una de las dos propiedades distribucionales y no alcanza el mínimo cuando a=0</p>
---	---

<p><b>Scott (1955)</b></p> $A_{62} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(b + c + 2d)}$	<p>Rango de variación [-1,1] No cumple una de las dos propiedades distribucionales y no alcanza el mínimo cuando a=0</p>
---	--

<p><b>Sokal &amp; Michener (1958), coeficiente de igualdad simple (SMC, por sus siglas en ingles <i>simple matching coefficient</i>)</b></p> $A_{64} = \frac{a + d}{n} = \frac{a + d}{a + b + c + d}$	<p>Rango de variación [0,1] No cumple la propiedad de alcanzar el mínimo cuando a=0</p>
---	---

<p><b>Sorensen (1948), Dice (1945), Gleason (1920), Czekanowski (1913) (coincide con la media armónica de Dice_1 y Dice_2)</b></p> $A_{70} = \frac{2a}{2a + b + c} = \frac{a}{a + \frac{1}{2}(b + c)}$	<p>Rango de variación [0,1] No cumple una de las dos propiedades distribucionales</p>
--	---

Los índices de Jaccard y Sorensen ya habían sido señalados como los más satisfactorios para un uso generalizado (Clifford & Stephenson 1975). Por ejemplo el segundo fue utilizado en estudios ecológicos marinos por Stephenson et al. 1970, donde fue erróneamente llamado Jaccard y por Gage (1972) donde fue por primera vez referido como Sorensen, ya que antes se lo conocía como Czekanowski.

Janson & Vegelius (1981) llaman al índice de Jaccard coeficiente de comunidad y afirman que puede considerarse uno de los más recomendables. Es monótono con respecto al índice de Sorensen, según Legendre & Legendre (1998) la diferencia entre ambos está dada únicamente por la escala.

Las conclusiones de este trabajo coinciden en parte con las encontradas por Warrens (2008), quien de la basta cantidad de coeficientes de similitud que figuran en su tesis bajo el título Lista de coeficientes de similitud, y de varias propiedades matemáticas estudiadas en algunos de ellos, afirma que se destacan siete de ellos, en el sentido de que para ellos se establecieron múltiples propiedades atractivas. Su conclusión es que para la mayoría de las aplicaciones analíticas de datos la elección para el coeficiente correcto para variables binarias puede probablemente estar limitada a los siguientes siete coeficientes: Braun-Blanquet, Jaccard, Sokal & Michener y Sorensen, Cohen, Loewinger y Russell & Rao. Las conclusiones de este trabajo coinciden con los cuatro primeros y a su vez seleccionamos otros distintos. Warrens (2008) hace incapié en las

propiedades de los índices cuando se generalizan como coeficientes multivariados, lo cual no es estudiado en este trabajo.

En esta tesis se encuentra que los índices de Cohen y Russell & Rao tienen buen comportamiento, solamente no cumplen con dos propiedades, pero una de ellas se refiere a los límites inferiores o superiores, que es obligatoria para Hubálek (1982) y en este trabajo se la considera una propiedad importante. El índice de Cohen no cumple con la Prop.3a), que especificaba que un índice de similitud, por ejemplo A, debería alcanzar su valor mínimo cuando  $a=d=0$ . Tampoco cumple con la Prop.5, la cual dijimos que es más estricta que la 3a), ya que exige alcanzar el valor mínimo cuando solamente  $a=0$  y no es tan importante. Por su parte el índice de Russell y Rao no cumple con la Prop.3b), que especificaba que un índice de similitud A debería alcanzar su valor máximo cuando  $b=c=0$ . La otra propiedad que no cumple es distribucional y no tan importante. Hubálek (1982) destacó la importancia de las primeras cinco propiedades para que un índice de similitud sea aceptable.

Por su parte el índice de Loevinger no cumple con varias propiedades, por ejemplo las Prop.3a) y 5 descritas recientemente, la Prop.1 que se refería a los límites que debían tener los índices (este varía entre  $[-\infty, 1]$ ), las propiedades distribucionales, la métrica, la euclideanidad y a su vez en el 30% de los casos estudiados en este trabajo se hace indeterminado.

Los índices que se consideran “los mejores” son aquellos que más propiedades cumplen y en especial, las propiedades consideradas por Hubálek (1982) como importantes y se refieren a los límites, la simetría, la monotonicidad, la linealidad, las distribuciones empíricas, el muestreo, la métrica, la euclideanidad y la indeterminancia.

Si se requiere una propiedad especial que cumpla un índice se puede recurrir a la Tabla 16 o a las Tablas 17, 18 y 19 para encontrar el mejor, por ejemplo, para Análisis Cluster, Gower & Legendre (1986) recomiendan utilizar índices con valores altos de Resolución, propiedad que está relacionada con la linealidad, y según este trabajo el índice con mayor Resolución es Fossum (12,919) y le sigue Harris & Lahey (10,228), ambos son convexos pero tienen un valor relativamente bajo de CV (32,58 y 34,10 respectivamente), por lo tanto pueden ser elegidos.

Si se trata de Análisis de ordenamiento, la linealidad parece principalmente deseable y además las propiedades métricas y euclidianas.

Entre los índices seleccionados tenemos algunos que no consideran las ausencias conjuntas y otros que sí, la elección entre unos y otros dependerá de la Ciencia que requiere el tipo de índice. Ya dijimos anteriormente que en Ecología de Comunidades son preferibles los primeros, es decir los que no consideran las ausencias conjuntas. Por su parte en Botánica, se vienen aplicando con éxito desde hace muchos años (Hubálek 1982) los segundos.

En el siguiente capítulo se encuentra una aplicación de los mejores índices a tres conjuntos de datos, una transecta real de datos de vegetación y una coenoclina y un coenoplano simulados, la primera respondiendo a un solo gradiente ambiental y la segunda a dos.

## CAPÍTULO 5

### EFICIENCIA DE LOS MEJORES ÍNDICES PARA EL ANÁLISIS DE DATOS DE VEGETACIÓN

En el Capítulo anterior se encontraron doce índices de similitud y sus tres disimilitudes derivadas que cumplen con la mayoría de las quince propiedades estudiadas en este trabajo. Los mismos se presentan como los mejores para un uso general, y responden a la pregunta de Hubálek (1982) acerca de cuáles son los mejores coeficientes de similitud. Sin embargo, este mismo autor señala que el problema de seleccionar índices debe ser resuelto en cada área de trabajo. Goodman & Kruskal (1954) ya señalaban que cada área científica que hace uso de medidas de asociación debería luego de estudios y ensayos, formalizar cuales de ellas son las más útiles para sus necesidades.

El área que nos interesa en este trabajo es Ecología de comunidades y en particular datos de presencia/ausencia de especies vegetales en unidades muestrales. Estas tablas de datos suelen presentar muchos ceros, a veces hasta un 80%, lo cual complica los análisis posteriores. Algunos autores (Pielou 1984, Orlóci 1978, Legendre & Legendre 1998) recomiendan con estos datos utilizar índices de similitud que no tengan en cuenta las ausencias conjuntas, ya que la ausencia de dos especies vegetales en un sitio no indica que las mismas están asociadas, como ya fue descrito en el Capítulo 1.

Los índices de similitud seleccionados como los mejores comprenden algunos que tienen en cuenta las ausencias conjuntas y otros que no. Por lo tanto resulta de interés probar cuál o cuáles de ellos se comportan mejor con datos de vegetación de la vida real o simulados.

El propósito de este Capítulo es aplicar los doce mejores índices de similitud seleccionados en este trabajo, a tres conjuntos de datos, una coenoclima y un coenoplano simulados y datos reales de vegetación. La coenoclima responde a un gradiente ambiental y el coenoplano a dos. Los mismos datos fueron utilizados en el trabajo de Camiz et al. (2017) para comparar distintas escalas para medir la vegetación.

### 5.1. Descripción de las bases de datos

Los conjuntos de datos sobre los que se aplican los índices corresponden a datos simulados y reales. Para la simulación se utilizan algoritmos que emulan las clásicas tablas de abundancia de las que se utilizan para describir la vegetación de zonas geográficas de interés. Como datos reales se consideran los resultados de una investigación realizada en el valle del arroyo Saladillo que han sido analizados en distintos contextos por la autora de esta tesis.

Los datos simulados fueron generados de acuerdo a Minchin (1987). La abundancia  $y$  de una especie en una unidad muestral de una comunidad dada, de acuerdo a un gradiente ecológico fue modelada a través de la distribución Beta, cuya expresión general responde a la función:

$$y = x^\alpha(1 - x)^\gamma \quad 0 \leq x \leq 1$$

donde  $y$  depende de la posición  $x_i$  de la unidad o punto de muestreo sobre el gradiente, normalizado para que esté comprendido en el intervalo  $[0,1]$ , y los dos parámetros  $\alpha$  y  $\gamma$  controlan la forma de la función. Minchin (1987) generaliza la función Beta para adaptarla a la simulación de respuestas unimodales de especies sobre cualquier rango de  $x$  a lo largo de un gradiente ambiental en términos de la abundancia de  $y$ . Esto está basado en  $\alpha$  y  $\gamma$ , en la posición de  $x$  en cada comunidad, en la posición  $m$  donde las especies tienen su modo o máximo  $M$ , y el rango  $r$  sobre el gradiente dentro del cual la abundancia es mayor que cero.

$$y = \begin{cases} \frac{M}{d} \left( \frac{x-m}{r} + b \right)^\alpha \left( 1 - \left( \frac{x-m}{r} + b \right) \right)^\gamma & m-br \leq x \leq r(1-b) \\ 0 & \text{en otro caso} \end{cases}$$

donde los dos parámetros  $b$  y  $d$  dependen solamente de  $\alpha$  y  $\gamma$  en la siguiente forma:

$$b = \frac{\alpha}{\alpha + \gamma}, \quad d = b^\alpha(1 - b)^\gamma$$

Para simular ruido, se suman números aleatorios a las abundancias generadas  $y$ . Para esta tarea se adoptó una distribución Binomial negativa. Esta distribución ha sido ampliamente usada para modelar abundancias de especies (White & Bennetts 1996, O'Hara & Kotze 2010, Pillar 2013), debido a que genera datos más similares a los datos reales de comunidad, con una gran cantidad de

números ceros (ausencias) cuando la abundancia esperada  $y$  es baja. Considerando un número  $\tilde{y}$  de sucesos en ensayos independientes con probabilidad de suceso fija igual a  $p$ , se extrajo un valor al azar desde la distribución Binomial negativa con un número de sucesos  $n = \tilde{y} * p / (1 - p)$ .

Las extracciones aleatorias de las distribuciones Beta generalizada y Binomial negativa fueron codificadas en lenguaje C++. Las versiones compiladas de los programas están disponibles en la siguiente página: <http://ecoqua.ecologia.ufrgs.br/arquivos/Software/DataSimulation>.

Agradecemos al Dr. Valério Pillar quién simuló las coenoclinas y cedió este material para utilizarlo en este trabajo.

Base 1: Coenoclina simulada correspondiente a un gradiente ambiental. La descripción de la coenoclina se realiza con una tabla con la abundancia de 60 especies en 30 unidades muestrales simuladas siguiendo un gradiente ambiental, para la cual se seleccionan al azar las ubicaciones de los puntos de muestreo ( $x_i$ ), la abundancia modal para cada especie ( $M$ ), la posición modal ( $m$ ), el rango ( $r$ ) (amplitud del nicho) y otros dos parámetros  $\alpha$  y  $\gamma$ , datos que intervienen en la identificación de la distribución Beta con la cual se generan los datos de abundancia.

La posición sobre el gradiente de cada unidad muestral ( $x_i$ ,  $i=1, \dots, 30$ ) se identifica a partir de un número aleatorio desde una distribución Uniforme entre 1 y 100. Es decir, los puntos de muestreo  $x_i$  se eligen al azar mediante esta selección. Sus coordenadas son luego estandarizadas para llevarlas a una escala (0,1).

Para cada especie se decide la abundancia modal  $M_j$ ,  $j=1, \dots, 60$  a partir de un número aleatorio desde una Uniforme entre 0 y 80 (este número interviene en la distribución beta que genera la abundancia).

Se elige aleatoriamente la posición modal  $m_j$ ,  $j=1, \dots, 60$  a partir de una Uniforme entre 0 y 100.

Se elige el rango  $r_j$ ,  $j=1, \dots, 60$  con un número aleatorio de una Uniforme entre 20 y 60.

Se eligen aleatoriamente los parámetros  $\alpha$  y  $\gamma$  desde una Uniforme entre 0,1 y 4. Por último se utiliza un modelo Beta que depende de  $M$ ,  $m$ ,  $r$ ,  $\alpha$  y  $\gamma$  para decidir sobre la abundancia de cada especie en cada punto de muestreo  $x_i$ , ( $y_j(x_i)$ ),

$i=1,\dots,30, j=1,\dots,60$ ). Se tienen las funciones de abundancia de las 60 especies y las coordenadas de las 30 unidades muestrales a lo largo de la coenoclina simulada (Figura 7).

Base 2: coenoplano simulado correspondiente a dos gradientes ambientales. Es una tabla con la abundancia de 60 especies en 30 unidades muestrales simuladas, basadas en dos gradientes ambientales. Todos los parámetros fueron generados como en la base 1, separadamente para cada gradiente, y 30 unidades muestrales simuladas fueron construidas, apareando aleatoriamente los valores de ambas coenoclinas. Como resultado de las simulaciones, para una de las especies resultaron todos los valores iguales a cero, por tal motivo, se consideraron 59 especies. Se tienen los pares de coordenadas de las 30 unidades muestrales en el coenoplano (Figura 8).

Base 3: una tabla de datos reales con 27 censos y 65 especies de la vegetación observada en un área de estudio situada en el valle del arroyo Saladillo, 10 km al sur de las localidades de Chabás y Sandford, pertenecientes al distrito Caseros de la provincia de Santa Fe. Cercano al río se encuentra una pradera de *Stipa hyalina* Nees (llamada *flechillar*), luego se ubican matas altas de *Spartina densiflora* Brongon (llamada *espartillar*) y más lejos del arroyo hay diferentes praderas halófilas (Carnevale et al. 1987, Carnevale & Torres 1990). Para tomar los datos de vegetación, se trazó una transecta perpendicular al arroyo Saladillo de 650 m, y se delimitaron cada 25 m, parcelas de 4 m<sup>2</sup>, sobre las que se identificaron todas las especies presentes junto con una estimación de la abundancia/cobertura según la escala combinada de Braun-Blanquet (1979).

Figura 7: funciones de abundancia de 60 especies simuladas a lo largo de un gradiente previo a la introducción de ruido y posición a lo largo del gradiente de 30 puntos de muestreo simulados aleatoriamente

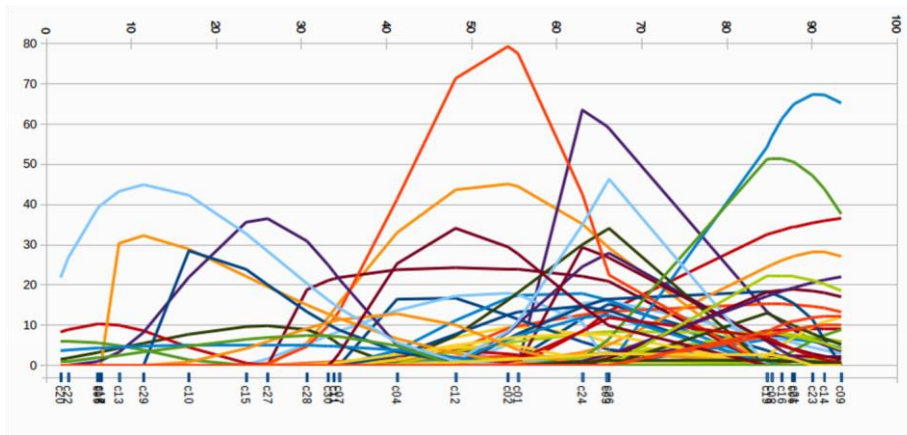
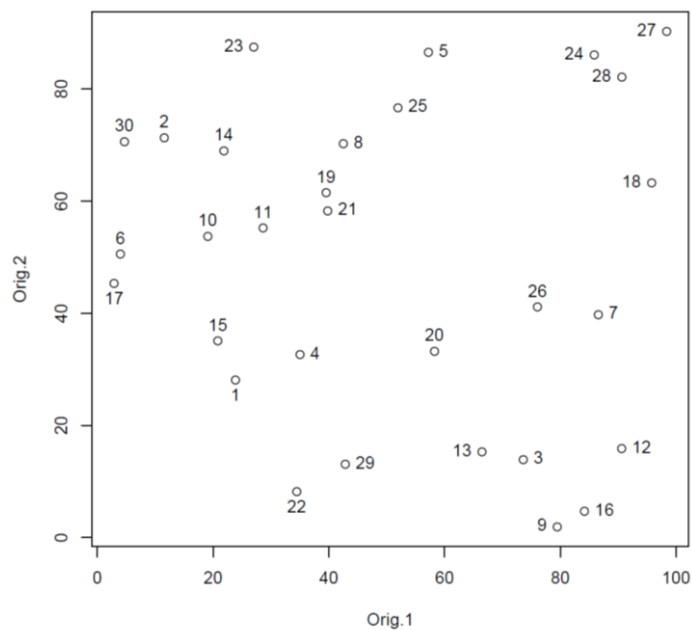


Figura 8: posición de 30 puntos de muestreo sobre el coenoplano resultante de los dos gradientes simulados



## 5. 2. Análisis multivariados sobre las bases de datos reales y simuladas

Las tres tablas de datos fueron convertidas a datos de presencia/ausencia, considerando la “presencia” para todo valor de abundancia distinto de cero. Para la coenoclima y el coenoplano, se calcularon para todos los pares de censos los doce índices de similitud considerados los “mejores” en este trabajo, obteniendo dos matrices de similitud de 30 x 30 para cada coeficiente. Para la tabla de datos

reales se procedió de la misma manera con los 27 censos, dando lugar a una matriz de similitud de 27 x 27 para cada coeficiente.

Los doce mejores índices de similitud y de las tres disimilitudes encontrados en este trabajo son los siguientes: Anónimo\_1, Anónimo\_2, Anónimo\_3, Braun-Blanquet, Hamann, Jaccard, Lamont & Grant, Maxwell & Pilliner, Rogot & Goldberg, Scott, Sokal & Michener y Sorensen.

Para trabajar las tres tablas de datos se pueden utilizar la similitud A o alguna de las tres disimilitudes estudiadas en este trabajo (D1, D2 y D3). Como el objetivo es aplicar a las mismas un Análisis de Coordenadas Principales (PCoA), se conoce que en general se toma como input una matriz de disimilitud, pero Gower (1966) ha mostrado que las coordenadas principales también pueden ser computadas directamente desde la similitud. La disimilitud 1 se descarta ya que no se encontraron propiedades euclídeanas para ninguno de los índices estudiados. Trabajar con la similitud A o D3 es equivalente si A es semi definida positiva y la similitud varía entre 0 y 1. Si esto no ocurre las representaciones que se obtienen aplicando PCoA con A y con D3 tienen distancias proporcionales, las cuales si se comparan con un método procrusteano son equivalentes. Lo mismo ocurre si se utiliza la disimilitud D2, por lo tanto se elige esta última como única medida de disimilitud.

A las tres matrices de similitud **D2** se les aplicó un Análisis de Coordenadas Principales (PCoA). Para comparar los comportamientos de los diferentes índices de disimilitud se utilizó Análisis Procrusteano Generalizado (GPA, Gower 1975, Gower & Dijksterhuis 2004, Camiz & Denimal 2011, Legendre & Legendre 2012, Lisboa et al. 2014). Las comparaciones fueron limitadas a las dos primeras coordenadas de los censos/unidades muestrales obtenidas de cada método, ya que como se espera solo uno o dos gradientes, son suficientes. GPA fue utilizado para chequear la homogeneidad de los resultados del mismo método de ordenamiento (PCoA) aplicado con diferentes índices de disimilitud. Básicamente GPA es la generalización del Análisis Procrusteano a más de dos configuraciones. Dadas dos nubes de puntos apareadas, el objetivo del Análisis procrusteano es encontrar la mejor transformación euclídea (traslación, escalamiento, reflexión y rotación) de una de las nubes para equipararse de la mejor manera posible a la otra, minimizando de esta forma la suma de las distancias al cuadrado de los pares de puntos. Generalizado a varias nubes de

puntos, se realiza un proceso iterativo que converge a un consenso, esto es una clase de promedio para cada punto, que mejor representa los distintos métodos. Esto permite comparar gráficamente las distintas nubes de puntos con el consenso. Desde GPA se obtiene la matriz de correlaciones de a pares procrusteanas, donde la correlación procrustea es un índice de asociación multidimensional de dos nubes de puntos. Este valor varía en el intervalo  $[0,1]$  y permite una comparación numérica de la representación de los censos/unidades muestrales obtenidos con diferentes métodos, ya que 1 significa que las nubes son idénticas y 0 que no tienen ninguna relación.

Para el cálculo de índices y la aplicación de PCoA se construyeron programas con el software IML de SAS. Se incluyen en el Apéndice 8 los programas 678 (página 330), 772 (página 335) y 882 (página 340) correspondientes a los datos reales, la coenoclina y el coenoplano respectivamente, todos para un índice tipo, en este caso Jaccard. Para la aplicación de GPA se utiliza el paquete FactoMineR del programa R (Husson et al. 2017).

#### Resultados Base 1:

Se presentan en la Tabla 21 las correlaciones procrusteanas entre las coordenadas de las 30 unidades muestrales tal como fueron simuladas para un único gradiente (llamadas X1), con respecto a las dos primeras coordenadas principales resultantes de los PCoA con diferentes índices. Se observa que la mayor correlación con las coordenadas originales se obtiene con los índices de Jaccard, Braun-Blanquet, Lamont & Grant y Sorensen, los cuales no contienen en su fórmula la cantidad "d". Un poco menor es la correlación con Rogot & Goldberg y Scott, que si consideran las ausencias conjuntas y de a poco va disminuyendo con los índices (Maxwell & Pilliner, Hamann, Sokal & Michener y Anónimo\_1). Por último están los índices Anónimo\_2 y Anónimo\_3 que tienen la menor correlación con las coordenadas originales y los mismos tienen la particularidad de considerar en su fórmula la cantidad "d" y no la cantidad "a". También destacamos las altas correlaciones entre todos los índices, y en particular se observan dos grupos muy correlacionados dentro y poco entre sí. El primero de ellos formado por Jaccard, Braun-Blanquet, **Lamont & Grant** y **Sorensen** (todos sin considerar la cantidad "d"), se destaca la igualdad entre los

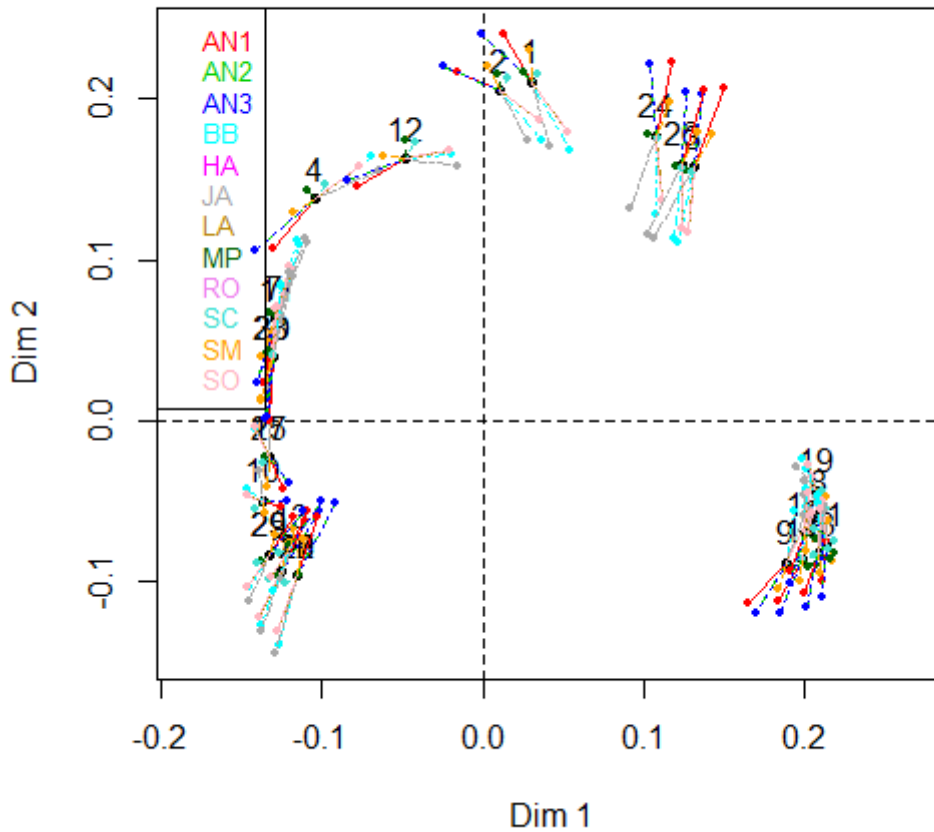
2 últimos índices. El segundo grupo está formado por **Rogot & Goldberg, Scott, Maxwell & Pilliner, Hamann, Sokal & Michener** y Anónimo\_1 (todos consideran la cantidad “d”), se destacan las igualdades entre los 2 primeros y los 2 penúltimos. Por su parte Anónimo\_2 y Anónimo\_3 presentan una correlación  $r=1$  entre sí y poca correlación con los restantes índices.

Tabla 21: Matriz de correlaciones procrusteanas entre las coordenadas de las 30 unidades muestrales tal como fueron simuladas para un único gradiente (llamadas X1), con respecto a las primeras coordenadas principales resultantes de los PCoA con diferentes índices de Similitud

	X1	JA	BB	LA	SO	RO	SC	MP	HA	SM	AN1	AN2	AN3
X1	1	0,9829	0,9764	0,9734	0,9734	0,9559	0,9559	0,9357	0,9283	0,9283	0,9032	0,8461	0,8461
JA		1	0,9983	0,9982	0,9982	0,9887	0,9887	0,9712	0,9686	0,9686	0,9435	0,8796	0,8796
BB			1	0,9995	0,9995	0,9853	0,9853	0,9636	0,9630	0,9630	0,9342	0,8608	0,8608
LA				1	1	0,9894	0,9894	0,9700	0,9698	0,9698	0,9427	0,8719	0,8719
SO					1	0,9894	0,9894	0,9700	0,9698	0,9698	0,9427	0,8719	0,8719
RO						1	1	0,9947	0,9947	0,9947	0,9808	0,9321	0,9321
SC							1	0,9947	0,9947	0,9947	0,9808	0,9321	0,9321
MP								1	0,9986	0,9986	0,9937	0,9634	0,9634
HA									1	1	0,9951	0,9597	0,9597
SM										1	0,9951	0,9597	0,9597
AN1											1	0,9810	0,9810
AN2												1	1
AN3													1

Se presenta a continuación la Figura 9 con los resultados del Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de disimilitud. Se observa para cada uno de los 30 puntos, su posición según los distintos métodos junto con su posición de compromiso encontrada por el método procrusteano. Se observa muy poca diferencia entre los métodos y con respecto al consenso, todos los métodos están bien correlacionados unos con otros, con correlaciones mayores a 0,84.

Figura 9: Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de Similitud para la primera coenoclima simulada.



#### Resultados Base 2:

Se presentan en la Tabla 22 las correlaciones procrusteanas entre las coordenadas de las 30 unidades muestrales tal como fueron simuladas para dos gradientes (llamadas X2), con respecto a las dos primeras coordenadas principales resultantes de los PCoA con diferentes índices. Se observa que la mayor correlación con las coordenadas originales se obtiene con los índices de Anónimo\_1, Lamont & Grant y Sorensen, de los cuales, el primero contiene en su formula la cantidad “d” mientras que los otros dos no. Un poco menor es la correlación con Maxwell & Pilliner que si considera las ausencias conjuntas y de

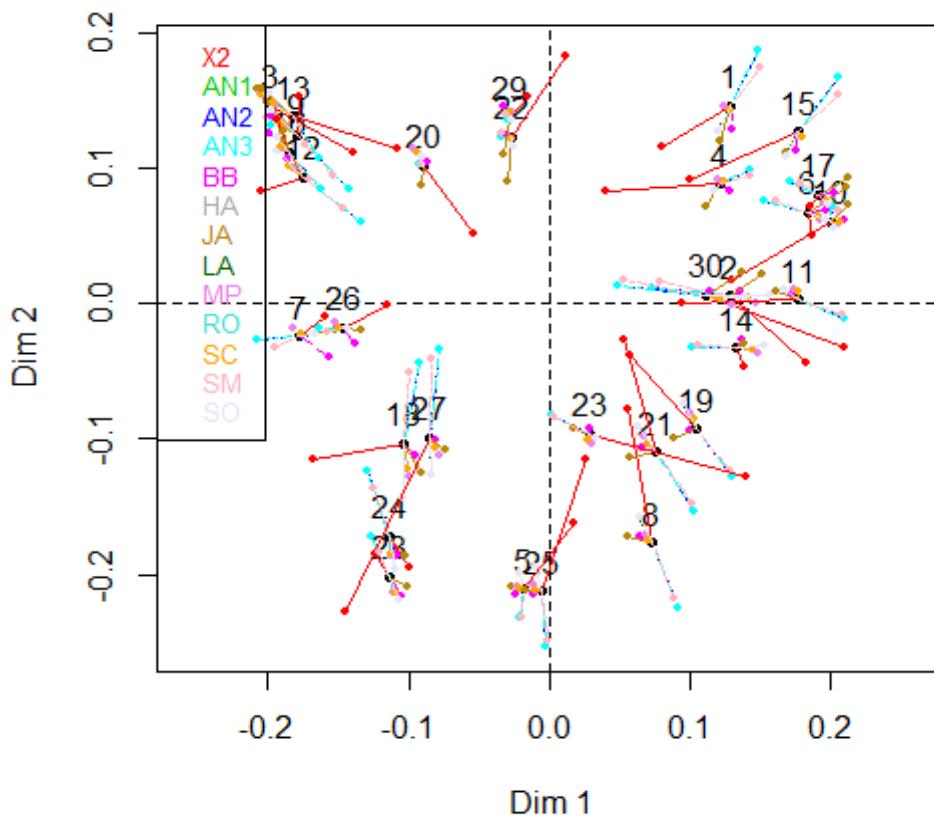
a poco va disminuyendo con los índices, alternando aquellos que consideran a “d” (Rogot & Goldberg, Scott, Hamann y Sokal & Michener) y aquellos que no (Braun-Blanquet y Jaccard). Por último están los índices Anónimo\_2 y Anónimo\_3 que tienen la menor correlación con las coordenadas originales y los mismos tienen la particularidad de considerar en su fórmula la cantidad “d” y no la cantidad “a”. También destacamos las altas correlaciones entre todos los índices, y en particular se observan dos grupos muy correlacionados dentro y poco entre sí. El primero de ellos formado por Anónimo\_1, Maxwell & Pilliner, Rogot & Goldberg, Scott, Braun-Blanquet, Jaccard, Lamont & Grant y Sorensen (los cuatro últimos no consideran la cantidad “d” mientras que los anteriores sí), se destaca la igualdad entre Anónimo\_1, Lamont & Grant y Sorensen por un lado y de Rogot & Goldberg y Scott por otro. El segundo grupo está formado por Hamann, Sokal & Michener, Anónimo\_2 y Anónimo\_3 (todos consideran la cantidad “d”), se destacan las igualdades entre los 2 primeros entre si y entre los 2 últimos entre sí.

Tabla 22: Matriz de correlaciones procrusteanas entre las coordenadas de las 30 unidades muestrales tal como fueron simuladas para dos gradientes (llamadas X2), con respecto a las primeras coordenadas principales resultantes de los PCoA con diferentes índices de Similitud

	X2	AN1	LA	SO	MP	RO	SC	BB	JA	HA	SM	AN2	AN3
X2	1	0,9388	0,9388	0,9388	0,9296	0,9257	0,9257	0,9215	0,9188	0,8527	0,8527	0,8372	0,8372
AN1		1	1	1	0,9982	0,9978	0,9978	0,9969	0,9957	0,9603	0,9603	0,9484	0,9484
LA			1	1	0,9982	0,9978	0,9978	0,9969	0,9957	0,9603	0,9603	0,9484	0,9484
SO				1	0,9982	0,9978	0,9978	0,9969	0,9957	0,9603	0,9603	0,9484	0,9484
MP					1	0,9997	0,9997	0,9976	0,9944	0,9731	0,9731	0,9639	0,9639
RO						1	1	0,9986	0,9951	0,9763	0,9763	0,9670	0,9670
SC							1	0,9986	0,9951	0,9763	0,9763	0,9670	0,9670
BB								1	0,9965	0,9744	0,9744	0,9638	0,9638
JA									1	0,9641	0,9641	0,9526	0,9526
HA										1	1	0,9988	0,9988
SM											1	0,9988	0,9988
AN2												1	1
AN3													1

Se presenta a continuación la Figura 10, con los resultados del Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de disimilitud y a las coordenadas originales del coenoplano simuladas siguiendo 2 gradientes. Se observa para cada uno de los 30 puntos, su posición según los distintos métodos junto con su posición de compromiso encontrada por el método procrusteano. Se observa muy poca diferencia entre los métodos y con respecto al consenso, existe alguna diferencia con respecto a las coordenadas originales. Todos los métodos están bien correlacionados unos con otros, con correlaciones mayores a 0,83.

Figura 10: Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de Similitud y a las coordenadas originales de la segunda coenoclina simulada.



### Resultados Base 3:

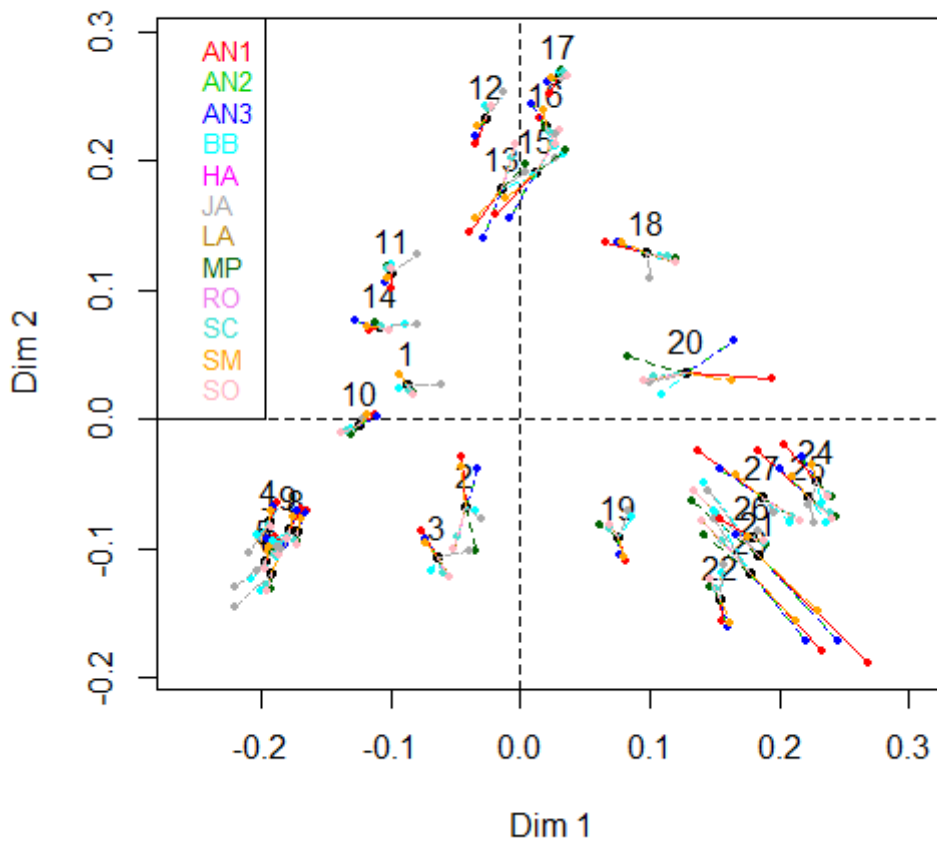
Se presentan en la Tabla 23 las correlaciones procrusteanas entre las dos primeras coordenadas principales resultantes de los PCoA aplicados con diferentes índices. No se tienen en este caso coordenadas originales de referencia, ya que se trata de datos reales de vegetación. Destacamos las altas correlaciones entre todos los índices, mayores a 0,93; y en particular se observan dos grupos muy correlacionados dentro y poco entre sí. El primero de ellos formado por Anónimo\_1, Anónimo\_2, Anónimo\_3, Hamann y Sokal & Michener (todos consideran la cantidad “d” en su fórmula), se destaca la igualdad entre Anónimo\_2 y Anónimo\_3 y entre los 2 últimos. El segundo grupo está formado por **Braun-Blanquet**, **Jaccard**, **Lamont & Grant**, Maxwell & Pilliner, Rogot & Goldberg, Scott y **Sorensen** (los que están en negritas no consideran la cantidad “d” mientras que el resto sí), se destacan las igualdades entre Lamont & Grant y Sorensen y entre Rogot & Goldberg y Scott.

Tabla 23: Matriz de correlaciones procrusteanas entre las coordenadas de las 27 unidades muestrales de la transecta de vegetación con respecto a las primeras coordenadas principales resultantes de los PCoA con diferentes índices de Similitud

	AN1	AN2	AN3	HA	SM	BB	JA	LA	MP	RO	SC	SO
AN1	1	0,9972	0,9972	0,9949	0,9949	0,9459	0,9443	0,9412	0,9433	0,9543	0,9543	0,9412
AN2		1	1	0,9982	0,9982	0,9602	0,9584	0,9563	0,9602	0,9679	0,9679	0,9563
AN3			1	0,9982	0,9982	0,9602	0,9584	0,9563	0,9602	0,9679	0,9679	0,9563
HA				1	1	0,9720	0,9697	0,9685	0,9704	0,9781	0,9781	0,9685
SM					1	0,9720	0,9697	0,9685	0,9704	0,9781	0,9781	0,9685
BB						1	0,9969	0,9975	0,9967	0,9977	0,9977	0,9975
JA							1	0,9949	0,9942	0,9951	0,9951	0,9949
LA								1	0,9989	0,9990	0,9990	1
MP									1	0,9989	0,9989	0,9989
RO										1	1	0,9990
SC											1	0,9990
SO												1

Se presenta a continuación la Figura 11, con los resultados del Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de disimilitud. Se observa para cada uno de los 30 puntos, su posición según los distintos métodos junto con su posición de compromiso encontrada por el método procrusteano. Se observa muy poca diferencia entre los métodos y con respecto al consenso, todos los métodos están bien correlacionados unos con otros, con correlaciones mayores a 0,94.

Figura 11: Análisis Procrusteano Generalizado aplicado a las 2 primeras coordenadas principales construidas con diferentes índices de Similitud para los datos de la transecta de vegetación.



Los resultados obtenidos corroboran las declaraciones de otros autores (Orlóci 1978, Pielou 1984, Legendre & Legendre 1998), quienes destacan el mejor desempeño de los índices de similitud que no consideran la cantidad “d” en su fórmula, cuando se analizan datos de vegetación. A tal efecto se recomiendan los índices de **Jaccard**, **Sorensen**, **Braun-Blanquet** y **Lamont & Grant**, sin embargo hay un índice que está muy correlacionado con los anteriores y es **Anónimo\_1**. Este último se lo atribuye a Van der Maarel (1969) en el paquete Simba de R, el mismo también parece resultar adecuado para datos de vegetación, si bien tiene en su fórmula la cantidad “d”. No sorprende encontrar que los índices de Jaccard y Sorensen son adecuados para analizar datos de comunidades vegetales, ya que al primero de ellos se lo conoce como coeficiente de comunidad (Legendre & Legendre 1998). También Janson & Vegelius (1981) y Warrens (2008) lo consideran uno de los coeficientes más recomendables. Por su parte Clifford & Stephenson (1975) recomiendan a Jaccard y Sorensen como los coeficientes más satisfactorios para un uso general. Además de estos 2 índices que han sido ampliamente utilizados en Ecología de comunidades, se descubren en esta tesis otros índices que también pueden dar resultados satisfactorios, como Anónimo\_1, Braun-Blanquet y Lamont & Grant.

## COMENTARIOS FINALES

Este trabajo se ha dedicado a realizar una extensa investigación comparativa entre un amplio conjunto de índices de similitud y disimilitud para datos binarios teniendo en cuenta distintas propiedades que han sido consideradas valiosas por varios autores a través de años de estudio.

La consideración amplia tanto de definiciones como de propiedades ha permitido deducir comportamientos sistemáticos que guiaron la selección de algunos indicadores como los mejores para un uso general. Estos índices satisfacen simultáneamente más de doce propiedades de las catorce estudiadas para las similitudes y tres o más de las seis estudiadas para las tres disimilitudes.

En particular, el interés de la autora está vinculado a los estudios de vegetación, por lo cual la tesis incorpora un estudio específico referido a esta área incluyendo el análisis multivariado de datos reales y de datos que simulan el comportamiento de la vegetación a lo largo de gradientes ambientales.

El uso de los indicadores seleccionados en estos escenarios dio por resultado un comportamiento consistente de la visualización de los parecidos entre puntos de muestreo dados por la aplicación de coordenadas principales a partir de distintas matrices de disimilitud, lo cual afirma la capacidad de estos indicadores seleccionados para identificar la asociación.

Por lo tanto el trabajo hace un aporte en el ámbito de los estudios multivariados, tanto de ordenamiento como de clasificación, los que requieren la transformación de la matriz de datos a una matriz de similitud o disimilitud, siendo la identificación de esta matriz un paso fundamental en los análisis ya que su definición condiciona todo resultado posterior, aspecto que es discutido por todos los autores.

En base a los resultados obtenidos en esta tesis, la elección de la medida de similitud/disimilitud puede fundarse en propiedades probadas objetivamente.

## BIBLIOGRAFÍA

Agrell I. 1945. An objective method for characterization of animal and plant communities. *K. fisiogr. Sällsk. Lund Förh.* 15: 63-67.

Anderson M.J. 2001a. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32-46.

Anderson M.J. 2001b. Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* 58: 626-639.

Austin B. & Colwell R.R. 1977. Evaluation of some coefficients for use in numerical taxonomy of micro-organisms. *International Journal of Systematic Bacteriology* 27: 204-210.

Barkman J.J. 1958. Phytosociology and Ecology of Cryptogamic Epiphytes, Including a Taxonomic Survey and Description of their Vegetation Units in Europe. Van Gorcum, Assen. pp.628.

Batagelj V. & Bren M. 1995. Comparing resemblance measures. *Journal of Classification* 12: 73-90.

Baroni-Urbani C. & Buser M.W. 1976. Similarity of binary data. *Systematic Zoology* 25: 251-259

Begon M., Harper J. L. & Townsend C. R. 1990. *ECOLOGY Individuals, Populations and Communities*. Blackwell Scientific Publications, London.

Benini R. 1901. Principii di Demografie. N° 29 of Manuali Barbéra di Science Giuridiche Sociali e Poltiche. Firenzi: G. Barbéra.

Braun-Blanquet J. (1928) *Pflanzensoziologir. Grundzuge der Vegetationskunde*. Springer. Wien, AT.

Braun-Blanquet J. (1932) *Plant Sociology: the Study of Plant Communities*. McGraw-Hills, New York.

Braun-Blanquet, J. (1979) *Fitosociología*. H. Blume Ediciones, Madrid

Bray J.R. 1956. A study of mutual occurrence of plant species. *Ecology* 37: 21-28.

Bray J.R. & Curtis J.T. 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* 27: 325-349.

Bren M. & Batagelj V. 2006. The metric index. *Croatica Chemica Acta* 79(3): 399-410.

Cailliez F. 1983. The analytical solution of the additive constant problem. *Psychometrika* 48, 2: 305-308.

Cain A.J. & Harrison G.A. 1958. An analysis of the taxonomist's judgment of affinity. *Proceedings of Zoological Society London* 131: 85-98.

Camiz S. 1999. Comparison of Euclidean approximations of non-Euclidean distances. In: *Classification and Data Analysis. Theory and Application*. M. Vichi & O. Opitz (eds.). Berlin, Springer. *Studies in Classification, Data Analysis and Knowledge Organization*, pp. 139-146.

Camiz S. & Denimal J.J. 2011. Procrustes analysis and stock markets. *Case Studies in Business, Industry and Government Statistics* 4: 93-100.

Camiz S., Torres P. & Pillar V.P. 2017. Recoding and multidimensional analyses of vegetation data: a comparison. *Community Ecology* 18(3): 260-279; 2017

Carnevale N. & Torres P. 1990. The relevance of physical factors on species distributions in inland salt marshes (Argentina). *Coenoses* 5: 113-120.

Carnevale N., Torres P., Boccanelli S. & Lewis J.P. 1987. Halophilous communities and species distributions along environmental gradients in southeastern Santa Fe Province, Argentina. *Coenoses* 2: 49-60.

Cattell R.B. 1952. Factor Analysis – an introduction and manual for the psychologist and social scientist. Harper, New York. pp.462.

Cheetham A.H. y Hazel J.E. 1969. Binary (presence-absence) similarity coefficients. *Journal of Paleontol.* 43: 1130-1136.

Clarke K.R. & Green R.H. 1988. Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Prog. Ser.* 46: 213-226.

Clarke K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18: 117-143.

Clement P.W. (1976). A formula for computing inter-observer agreement. *Psychological Reports* 39: 257-258.

Clifford H.T. & W. Stephenson. 1975. An Introduction to Numerical Classification. Academic Press. NY, USA. pp.229.

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37-46.

Cole L.C. 1949. The measurement of interspecific association. *Ecology* 30: 411-424.

Cole L.C. 1957. The measurement of partial interspecific association. *Ecology* 38: 226-233.

Cox T.F. & Cox M.A.A. 2000. A general weighted two-way dissimilarity coefficient. *Journal of Classification* 17: 101-121.

Cronbach L.J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297-334.

Cuadras C.M. 1989. *Estadística Española*, Vol. 30, Núm. 119, págs. 295 a 378.

Cuadras C.M. 1996. *Métodos de análisis multivariante*. EUB. Ed. SL. Barcelona. 642pp.

Cullen A. & Frey H. 1999. *Probabilistic Techniques in Exposure Assessment*. First edition. Plenum Publishing Co., New York.

Czekanowski J. 1913. Zarys Metod Statystycznych w Zastosowaniu do Antropologii. *Travaux de la Société des Sciences de Varsoie*. III. Classe des sciences mathématiques et naturelles, núm. 5. iv + 228pp.

Czekanowski J. 1932. Coefficient of racial likeness und Durchschnittliche Differenz. *Anthropologischer Anzeiger* 9: 227-249.

Dagnelie P. 1960. Contribution á l'étude des communautés végétales par l'analyse factorielle. *Bull. Serv. Carte Phytogéogr.* B 5: 7-71, 93-195.

Delignette-Muller M., Pouillot R., Denis J., Dutang C. 2013. *fitdistrplus: Help to Fit a Parametric Distribution to Non-Censored or Censored Data*. R package version 1.0-1, URL. <http://cran.r-project.org/web/packages/fitdistrplus/>.

Demey J.R., Pla L., Vicente-Villardón J.L., Di Rienzo J.A., Casanoves F. 2011. Medidas de distancia y de similitud. En *Valoración y análisis de la diversidad funcional y su relación con los servicios ecosistémicos*. Editores: F. Casanoves, L. Pla y J.A. Di Rienzo. Serie Técnica N° 384. CATIE. Turrialba, Costa Rica. pp.47-59.

Dice L.R. (1945). Measures of the amount of ecological association between species. *Ecology* 26: 297-302.

Digby P.G.N. 1983. Approximating the tetrachoric correlation coefficient. *Biometrics* 39: 753-757.

Doolittle M.H. 1885. The verification of predictions. *Bulletin of the Philosophical Society of Washington* 7: 122-127.

Driver H.E. & Kroeber A.L. 1932. Quantitative expression of cultural relationships. *The University of California Publications in American Archaeology and Ethnology* 31: 211-256.

Edgington E.S. 1987. Randomization tests. Marcel Dekker. New York.

Efron, B. & Tibshirani, R. 1994. An introduction to the Bootstrap. Chapman & Hall, New York.

Ellis D., Furner-Hines J., Willett P. 1993. Measuring the degree of similarity between objects in text retrieval systems. *Perspectives in Information Management* 3(2): 128-149.

Eyraud H. 1936. Les principes de la mesure des correlations. *Annales de l'Université de Lyon. Série III, Section A, 1*: 30-47.

Faith D.P. 1983. Asymmetric binary similarity measures. *Oecologia* 57: 287-290.

Fager E.W. & McGowan J.A. 1963. Zooplankton species groups in the North Pacific. *Science* 140: 453-460.

Fichet B. y Le Calve G. 1984. Structure geometrique des principaux indices de dissimilarité sur signes de presence-absence. *Statistiques et Analyse de dones* Vol. 9 N° 3: 11-44.

Fleiss J.L. 1975. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics* 31: 651-659.

Florek K., Lukaszewics J., Perkal J., Steinhaus H., Zubrzycki S. 1951. Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum* 2: 282–285.

Forbes S.A. 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. *Bulletin of the Illinois State Laboratory for Natural History* 7: 273-303.

Forbes S.A. 1925. Method of determining and measuring the associative relations of species. *Science* 61: 524.

Gage J. 1972. A preliminary survey of the benthic macrofauna and sediments in Lochs Etive and Creran, sea-lochs along the west coast of Scotland. *J. Mar. Biol. Ass. U.K.* 52: 237-276.

Giansanti C. 1990. Rappresentazioni geometriche indotte da coefficienti di associazione. Tesi di Laurea in Matematica. Facoltà di Scienze Matematiche, Fisiche e Naturali. Università degli studi di Roma "La Sapienza".

Gilbert N. & Wells T.C.E. 1966. Analysis of quadrat data. *Journal of Ecology* 54: 675-685.

Gleason H.A. 1920. Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club* 47: 21-33.

Goodall D.W. 1954. Vegetation classification and vegetation continua. *Angew. Pflanzensoz. Viena. Festschr. Aichinger* 1: 168-182.

Goodall D.W. 1973. Sample similarity and species correlations. In *Ordination and Classifications of Communities* (ed. R.H. Whittaker). W. Junk, The Hague. pp.105-156.

Goodall D.W. 1967. The distribution of the matching coefficients. *Biometrics* 23: 647-656.

Goodman L.A. y Kruskal W.H. 1954. Measures of association for cross classifications I. *Journal of the American Statistical Association* 49: 732-764.

Goodman L.A. y Kruskal W.H. 1959. Measures of association for cross classifications II. Further discussion and references. *Journal of the American Statistical Association* 54: 123-163.

Goodman L.A. y Kruskal W.H. 1963 Measures of association for cross classifications III. Approximate sampling theory. *Journal of the American Statistical Association* 58: 310-364.

Gower J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325-338.

Gower J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-874.

Gower J.C. 1975. Generalized Procrustes analysis. *Psychometrika* 40: 33-51.

Gower J.C. 1982. Euclidean Distance Geometry. *Mathematical Scientist* 7: 1-14.

Gower J.C. 1985. Properties of Euclidean and Non-Euclidean Distance Matrices. *Linear Algebra and its Applications* 67: 81-95.

Gower J.C. & Dijksterhuis G.B. 2004. *Procrustes Problems*. Oxford University Press, Oxford.

Gower J.C. y Legendre P. 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* 3: 5-48.

Hajdu L.J. 1981. Graphical Comparison of Resemblance Measures in Phytosociology. *Vegetatio* 48: 47-59.

Hamann U. 1961. Merkmalsbestand und Verwandtschaftsbeziehungen der Farinose. Ein Betrag zum System der Monokotyledonen. *Willdenowia* 2: 639-768.

Harris F.C. & Lahey B.B. 1978. A method for combining occurrence and nonoccurrence agreement scores. *Journal of Applied Behavioral Analysis* 11: 523-527.

Hawkins R.P. & Dotson V.A. 1968. Reliability scores that delude: An Alice in Wonderland trip through the misleading characteristics of interobserver agreement scores in interval coding. In: E. Ramp & G. Semb (Eds.), *Behavior Analysis: Areas of Research and Application*. Englewood Cliffs, N.J., Prentice-Hall. pp.539-576.

Holley J.W. & Guilford J.P. 1964. A note on the G-index of agreement. *Educational and Psychological Measurement* 24: 749-753.

Holliday J.D., Hu C.Y. & Willett P. 2002. Grouping of Coefficients for the Calculation of Inter-Molecular Similarity and Dissimilarity using 2D Fragment Bit-Strings. *Combinatorial Chemistry & High Throughput Screening* 5: 155-166.

Hubálek Z. 1974. Dispersal of fungi of the family Chaetomiaceae by free-living birds. I. A survey of records. *Ceská Mykologie* 28: 65-79.

Hubálek Z. 1978. Coincidence of fungal species associated with birds. *Ecology* 59: 438-442.

Hubálek Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biolog. Rev.* 57: 669-689.

Hubert L.J. 1977. Nominal scale response agreement as a generalized correlation. *British Journal of Mathematical and Statistical Psychology* 30: 98-103.

Hurlbert S.H. 1969. A coefficient of interspecific association. *Ecology* 50: 1-9.

Husson F., Lé S. & Pagés J. 2017. Exploratory Multivariate Analysis by Example Using R. CRC Press, Boca Raton, FL.

Iversen J. 1954. Über die Korrelationen zwischen den Pflanzenarten in einem grönländischen Talgebiet. *Vegetatio* 5-6: 238-246.

Jaccard P. 1901. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37: 241-272.

Janowitz M.F. 1980. Similarity measures on binary data. *Systematic Zoology* 29: 342-359.

Janson S. & Vegelius J. 1981. Measures of ecological association. *Oecologia* 49: 371-376.

Jardine N. & Sibson R. 1971. *Mathematical Taxonomy*. John Wiley and Sons, London-New York-Sydney-Toronto.

Johnson S.C. 1967. Hierarchical clustering schemes. *Psychometrika* 32: 241-254.

Johnson J.G. 1971. A quantitative approach to faunal province analysis. *American Journal of Science* 270: 257-280.

Jurasinski G. 2012. Package 'simba'. A collection of functions for similarity analysis of vegetation data.

Kendall M.G. & Stuart A. 1967. *The Advanced Theory of Statistics*. 2nd. ed. Charles Griffin. London

Kenkel N.C. y Booth T. 1987. A comparison of presence-absence resemblance coefficients for use in biogeographical studies. *Coenoses* 2(1): 25-30.

Koleff P., Gaston K.J. & Lennon J.J. 2003. Measuring beta diversity for presence-absence data. *Journal of Animal Ecology* 72: 367-382.

Kuder G.F. & Richardson M.W. 1937. The theory of estimation of test reliability. *Psychometrika* 2: 151-160.

Kulczynski S. 1927. Die Pflanzenassoziationen der Pienenen. Bulletin International de l'Académie Polonaise des Sciences et des Letters, Classe des Sciences Mathématiques et Naturelles, Serie B, Supplément II, 2: 57-203.

Lamont B.B. & Grant K.J. 1979. A comparison of twenty-one measures of site dissimilarity. In: Orlóci L., Rao C.R. & Stiteler W.M. (eds.). *Multivariate Methods in Ecological Work*, pp. 101-126. International Co-operative Publishing House, Fairland, MD. 550pp.

Legendre L. y Legendre P. 1998. *Numerical ecology*. Elsevier Scientific Publishing Company, Amsterdam, 853pp

Legendre P. y Legendre L.F.J. 2012. *Numerical ecology*. 3rd.ed. Elsevier Scientific Publishing Company, Amsterdam

Lesica P. & B.M. Steele. 1997. Use of permanent plots in monitoring plant populations. *Natural Areas Journal* 17: 331-340.

Lingoes J.C. 1971. Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika* 36: 195-203.

Lisboa F.J.G., Peres-Neto P.R., Chaer G.M., Da Conceio Jesus E., Mitchell R.J., Chapman S.J. & Berbara R.L.L. 2014. Much beyond Mantel: bringing Procrustes association metric to the plant and soil ecologist's toolbox. *PLoS ONE* 9(6):e101238.

Loevinger J.A. 1947. A systematic approach to the construction and evaluation of tests of ability. *Psychometrika*, Monograph N° 4.

Loevinger J.A. 1948. The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin* 45: 507-530.

Manly B.F.J. 1998. Randomization, bootstrap and Monte Carlo methods in Biology. Chapman & Hall, London, UK.

Mardia K.V. 1978. Some properties of classical multi-dimensional scaling. Comm. in Statistics – Theory and Methods. Series A, 7, 13: 1233-1241.

Maxwell A.E. & Pilliner A.E.G. 1968. Deriving coefficients of reliability and agreement for ratings. *British Journal of Mathematical and Statistical Psychology* 21: 105-116.

Mc Connaughey B.H. 1964. The determination and analysis of plankton communities. Marine Research, Special N°, Indonesia, 1-40.

McCune, B. and J.B. Grace. 2002. Analysis of Ecological Communities. MjM Software Design. Gleneden Beach, Oregon. USA. pp.300.

McCune, B. and M. J. Mefford. 2016. PC-ORD. Multivariate Analysis of Ecological Data. Version 7.02. MjM Software, Gleneden Beach, Oregon, U.S.A.

Michael E.L. 1920. Marine ecology and the coefficient of association. *Journal of Animal Ecology* 8: 54-59.

Minchin P.R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71: 145-156.

Moore A.W. & Russell J.S. 1967. Comparison of coefficients and grouping procedures in numerical analysis of soil trace element data. *Geoderma* 1: 139-158.

Mountford M.D. 1962. An index of similarity and its applications to classificatory problems. In: P.W. Murphy (Ed.), *Progress in Soil Zoology* London. Butterworths. pp.43-50.

Nei M. & Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76: 5269-5273.

O'Hara R.B. & Kotze D.J. 2010. Do not log-transform count data. *Meth. Ecol. Evol.* 1: 118-122.

Ochiai A. 1957. Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Bulletin of the Japanese Society for Fish Science* 22: 526-530.

Odum E.P. 1950. Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology* 31: 587-605.

Orlóci L. 1978. *Multivariate analysis in vegetation research*. Dr. W.Junk b.v. The Hague. pp.451.

Pearson K. 1926. On the coefficient of racial likeness. *Biometrika* 18: 105-117.

Pearson K. & Heron D. 1913. On theories of association. *Biometrika* 9: 159-315.

Peirce C.S. 1884. The numerical measure of the success of predictions. *Science* 4: 453-454.

Peck J. E. 2010. *Multivariate Analysis for Community Ecologists: Step-by-Step using PC-ORD*. MjM Software Design. Gleneden Beach, OR. 162p.

Pielou E.C. 1969. *An Introduction to Mathematical Ecology*. Wiley-Interscience. New York.

Pielou E.C. 1984. *The Interpretation of Ecological Data. A primer on classification and ordination*. John Wiley & Sons. NY, USA. pp.263.

Pillar V.D.P. & Orlóci L. 1996. On randomization testing in vegetation science: multifactor comparisons of relevé groups. *J. Veg. Sc.* 7: 585-592.

Pillar V.D. 2013. How accurate and powerful are randomization tests in multivariate analysis of variance? *Community Ecol.* 14: 153-163.

Podani J. 1997. A measure of discordance for partially ranked data when presence/absence is also meaningful. *Coenoses* 12(2-3): 127-130.

Rodríguez-Salazar M.E., Álvarez-Hernández S. & Bravo-Núñez E. 2001. *Coeficientes de asociación*. Plaza y Valdés Editores. México.

Rogers D.J. & Tanimoto T.T. 1960. A computer program for classifying plants. *Science* 132: 1115-1118.

Rogot E. & Goldberg I.D. 1966. A proposed index for measuring agreement in test-retest studies. *Journal of Chronic Disease* 19: 991-1006.

Ruggiero A., Lawton J.H. & Blackburn T.M. 1998. The geographic ranges of mammalian species in South America: spatial patterns in environmental resistance and anisotropy. *Journal of Biogeography* 25: 1093-1103.

Russell P.F. & Rao T.R. 1940. On habitat and association of species of Anopheline larvae in South-Eastern Madras. *Journal of Malaria Institute India* 3: 153-178.

Scott W.A. 1955. Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly* 19: 321-325.

Shi G.R. 1993. Multivariate data analysis in palaeoecology and palaeobiogeography-a review. *Palaeogeography, Palaeoclimatology, Palaeoecology* 105: 199-234.

Simpson G.G. 1943. Mammals and nature of continents. *American Journal of Science* 241: 1-31.

Simpson G.G. 1960. Notes on the measurement of faunal resemblance. *American Journal of Science, Bradley volume*, 258-A: 300-311.

Sneath P.H.A. 1957a. Some thoughts on bacterial classification. *Journal of General Microbiology* 17: 184-200.

Sneath P.H.A. 1957b. The application of computers to taxonomy. *Journal of General Microbiology* 17: 201-226.

Sneath P.H.A. & Sokal R.R. 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco.

Sokal R.R. & Michener C.D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38: 1409-1438.

Sokal R.R. & Sneath P.H. 1963. Principles of Numerical Taxonomy. W.H. Freeman & Co. San Francisco.

Sorensen T. 1948. A method of stabilizing groups of equivalent amplitude in plant sociology based on the similarity of species content and its application to

analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab Biologiske Skrifter* 5: 1-34.

Sorgenfrei T. 1959. Molluscan assemblages from the marine middle Miocene of South Jutland and their environments. *Danmark Geologiske Undersøgelse. Serie 2*, 79: 403-408.

SAS Institute Inc. 2004. SAS/IML 9.1 User's Guide. Cary, NC: SAS Institute Inc.

Stephenson W., Williams W.T. & Lance G.N. 1970. The macrobenthos of Moreton Bay. *Ecol. Monogr.* 40: 459-494.

Stiles H.E. 1961. The association factor in information retrieval. *Journal of the Association for Computing Machinery* 8: 271-279.

Tarwid K. 1960. Szacowanie zbiezności nisz ekologicznych gatunków droga oceny prawdopodobieństwa spotykania się ich w polowach. *Ekologia Polska, Ser. B*, 6: 115-130.

Torgerson W.S. 1958. *Theory and Methods of Scaling*. J.Wiley & Sons., New York.

Torres P.S., Quaglino M.B. & Pillar V.D. 2010. Properties of a randomization test for multifactor comparisons of groups. *Journal of Statistical Computation and Simulation* 80: 1131-1150.

T'Mannetje L. 1967. A comparison of eight numerical procedures applied to the classification of some African *Trifolium* taxa based on *Rhizobium* affinities. *Australian Journal of Botany* 15: 521-528.

Van der Maarel E. 1966. Over vegetatiestructuren, -relaties en -systemen in het bijzonder in de duingraslanden van Voorne. Thesis, Univ. Utrecht. pp.170

Van der Maarel E. 1969. On the use of ordination models in phytosociology. *Vegetatio* 19: 21-46.

Warrens M.J. 2008. Similarity Coefficients for Binary Data. Properties of coefficients, coefficients matrices, multi-way metrics and multivariate coefficients. Dissertation Leiden University. pp.237.

White G.C. & Bennetts R.E. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology* 77: 2549-2557.

Wildi O. 2010. Data Analysis in Vegetation Ecology. John Wiley & Sons. UK. pp.211.

Williams W.T. y Dale M.B. 1965. Fundamental problems in numerical taxonomy. *Advanced Bot. Res.* 2: 35-68.

Williams W.T., Lambert J.M. & Lance G.N. 1966. Multivariate methods in plant ecology. V. Similarity analyses and information-analysis. *Journal of Ecology* 54: 427-445.

Yule G.U. 1900. On the association of attributes in statistics. *Philosophical Transactions of the Royal Society A*, 194: 257-319.

Yule G.U. 1912. On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society* 75: 579-652.

Yule G.U. & Kendall M.G. 1950. An Introduction to the Theory of Statistics, 14<sup>th</sup> ed. Charles Griffin and Co. Ltd., London.

## ANEXO 1

Propiedad 3a: Existencia de valores mínimos

$$\text{mín } A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a=0,b,c,d=0)}$$

**Anónimo\_1** [-1, 1]

$$A_1(j, k)^{(a,b,c,d)} = \frac{2a - (b + c)}{2a + (b + c)}$$

$$A_1(j, k)^{(a=0,b,c,d=0)} = \frac{2 * 0 - (b + c)}{2 * 0 + (b + c)} = \frac{-(b + c)}{(b + c)} = -1 \text{ (mínimo)}$$

**Anónimo\_2** [0, 1]

$$A_2(j, k)^{(a,b,c,d)} = \frac{2d}{2d + b + c}$$

$$A_2(j, k)^{(a=0,b,c,d=0)} = \frac{2 * 0}{2 * 0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Anónimo\_3** [-1, 1]

$$A_3(j, k)^{(a,b,c,d)} = \frac{2d - (b + c)}{2d + (b + c)}$$

$$A_3(j, k)^{(a=0,b,c,d=0)} = \frac{2 * 0 - (b + c)}{2 * 0 + (b + c)} = \frac{-(b + c)}{(b + c)} = \frac{-(b + c)}{(b + c)} = -1 \text{ (mínimo)}$$

**Anónimo\_4** [0, 1]

$$A_4(j, k)^{(a,b,c,d)} = \frac{4ad}{4ad + (a + d)(b + c)}$$

$$A_4(j, k)^{(a=0,b,c,d=0)} = \frac{4 * 0 * 0}{4 * 0 * 0 + (0 + 0)(b + c)} = \frac{0}{0 + 0} \\ = \text{INDETERMINADO, NO ES EL MÍNIMO}$$

**Anónimo\_5** [-1, +∞]

$$A_5(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a + b)(a + c), (c + d)(b + d)\}}$$

$$A_5(j, k)^{(a=0,b,c,d=0)} = \frac{0 * 0 - bc}{\text{mín}\{(0 + b)(0 + c), (c + 0)(b + 0)\}} = \frac{-bc}{\text{mín}\{(bc), (cb)\}} = \frac{-bc}{bc} \\ = -1 \text{ (mínimo)}$$

**Baroni-Urbani & Buser\_1** [0, 1]

$$A_6(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

$$A_6(j, k)^{(a=0,b,c,d=0)} = \frac{\sqrt{0 * 0} + 0}{\sqrt{0 * 0} + 0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Baroni-Urbani & Buser\_2** [-1, 1]

$$A_7(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a - b - c}{\sqrt{ad} + a + b + c}$$

$$A_7(j, k)^{(a=0,b,c,d=0)} = \frac{\sqrt{0 * 0} + 0 - b - c}{\sqrt{0 * 0} + 0 + b + c} = \frac{-b - c}{b + c} = \frac{-(b + c)}{b + c} = -1 \text{ (mínimo)}$$

**Braun-Blanquet** [0, 1]

$$A_8(j, k)^{(a,b,c,d)} = \frac{a}{\text{máx.}\{(a + b), (a + c)\}}$$

$$A_8(j, k)^{(a=0,b,c,d=0)} = \frac{0}{\text{máx.}\{(0 + b), (0 + c)\}} = \frac{0}{\text{máx.}\{(b), (c)\}} = 0 \text{ (mínimo)}$$

**Clement** [0, n]

$$A_9(j, k)^{(a,b,c,d)} = \frac{a(c + d)}{(a + b)} + \frac{d(a + b)}{(c + d)}$$

$$A_9(j, k)^{(a=0,b,c,d=0)} = \frac{0(c + 0)}{(0 + b)} + \frac{0(0 + b)}{(c + 0)} = \frac{0}{b} + \frac{0}{c} = 0 \text{ (mínimo)}$$

**Clifford & Stephenson** [0, 1]

$$A_{10}(j, k)^{(a,b,c,d)} = \frac{\chi^2}{n} = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{10}(j, k)^{(a=0,b,c,d=0)} = \frac{(00 - bc)^2}{(0 + b)(c + 0)(0 + c)(b + 0)} = \frac{(-bc)^2}{bc cb} = \frac{(bc)^2}{(bc)^2} = 1 \text{ NO ES EL MÍNIMO, ES EL MÁXIMO}$$

**Cohen** [-1, 1]

$$A_{11}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}$$

$$A_{11}(j, k)^{(a=0, b, c, d=0)} = \frac{2(0 \cdot 0 - bc)}{(0 + b)(b + 0) + (0 + c)(c + 0)} = \frac{-2bc}{bb + cc}$$

$$= \frac{-2bc}{b^2 + c^2} \quad \text{NO ES EL MÍNIMO}$$

si  $b = c \rightarrow A_{11}(j, k)^{(a=0, b, c, d=0)} = \frac{-2cc}{c^2 + c^2} = \frac{-2c^2}{2c^2} = -1$  (mínimo)

**Cole\_1** [-1, 1]

$$A_{12}(j, k)^{(a, b, c, d)} = \frac{\sqrt{2}(ad - bc)}{\sqrt{(ad - bc)^2 + (a + b)(c + d)(a + c)(b + d)}}$$

$$A_{12}(j, k)^{(a=0, b, c, d=0)} = \frac{\sqrt{2}(0 \cdot 0 - bc)}{\sqrt{(0 \cdot 0 - bc)^2 + (0 + b)(c + 0)(0 + c)(b + 0)}} = \frac{-\sqrt{2}bc}{\sqrt{(-bc)^2 + bccb}}$$

$$= \frac{-\sqrt{2}bc}{\sqrt{(-bc)^2 + (bc)^2}} = \frac{-\sqrt{2}bc}{\sqrt{2(bc)^2}} = \frac{-\sqrt{2}bc}{\sqrt{2}bc} = -1$$
 (mínimo)

**Cole\_2** [0, 1]

$$A_{13}(j, k)^{(a, b, c, d)} = \frac{\Phi}{\Phi_{\text{máx.}}} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} \cdot \frac{\Phi_{\text{máx.}}}{\Phi_{\text{máx.}}}$$

$$A_{13}(j, k)^{(a=0, b, c, d=0)} = \frac{0 \cdot 0 - bc}{\sqrt{(0 + b)(c + 0)(0 + c)(b + 0)}} = \frac{-bc}{\sqrt{(bc)^2}} = \frac{-1}{\Phi_{\text{máx.}}}$$

Si  $\Phi_{\text{máx}} = 1 \rightarrow A_{13}(j, k)^{(a=0, b, c, d=0)} = -1$  NO ES EL MÍNIMO

**Cole\_3**  $[-\infty, 1]$

$$A_{14}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a + b)(b + d)}$$

$$A_{14}(j, k)^{(a=0, b, c, d=0)} = \frac{0 \cdot 0 - bc}{(0 + b)(b + 0)} = \frac{-bc}{bb} = \frac{-c}{b} \quad \text{NO ES EL MÍNIMO}$$

si  $b = c \rightarrow A_{14}(j, k)^{(a=0, b, c, d=0)} = \frac{-c}{c} = -1$  (mínimo)

**Cole\_4**  $[-\infty, 1]$

$$A_{15}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a + c)(c + d)}$$

$$A_{15}(j, k)^{(a=0, b, c, d=0)} = \frac{0 \cdot 0 - bc}{(0 + c)(c + 0)} = \frac{-bc}{cc} = \frac{-b}{c} \quad \text{NO ES EL MÍNIMO}$$

$$\text{si } b = c \rightarrow A_{15}(j, k)^{(a=0, b, c, d=0)} = \frac{-c}{c} = -1 \text{ (mínimo)}$$

**Dennis**  $[-\infty, +\infty]$

$$A_{16}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{\sqrt{(a + b + c + d)(a + b)(a + c)}}$$

$$A_{16}(j, k)^{(a=0, b, c, d=0)} = \frac{0 \cdot 0 - bc}{\sqrt{(0 + b + c + 0)(0 + b)(0 + c)}} = \frac{-bc}{\sqrt{(b + c)bc}} \text{ NO ES EL MÍNIMO}$$

$$\text{si } b = c \rightarrow A_{16}(j, k)^{(a=0, b, c, d=0)} = \frac{-cc}{\sqrt{(c + c)cc}} = \frac{-c^2}{\sqrt{(2c)c^2}} = \frac{-1}{c\sqrt{2c}} < -1 \text{ NO ES EL MÍNIMO}$$

**Dice\_1**  $[0, 1]$

$$A_{17}(j, k)^{(a, b, c, d)} = \frac{a}{a + b}$$

$$A_{17}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{0 + b} = \frac{0}{b} = 0 \text{ (mínimo)}$$

**Dice\_2**  $[0, 1]$

$$A_{18}(j, k)^{(a, b, c, d)} = \frac{a}{a + c}$$

$$A_{18}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{0 + c} = \frac{0}{c} = 0 \text{ (mínimo)}$$

**Dice\_3**  $[0, 2]$

$$A_{19}(j, k)^{(a, b, c, d)} = \frac{2a}{(a + b)(a + c)}$$

$$A_{19}(j, k)^{(a=0, b, c, d=0)} = \frac{2 * 0}{(0 + b)(0 + c)} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Digby**  $[-1, 1]$

$$A_{20}(j, k)^{(a, b, c, d)} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}}$$

$$A_{20}(j, k)^{(a=0, b, c, d=0)} = \frac{(0 * 0)^{3/4} - (bc)^{3/4}}{(0 * 0)^{3/4} + (bc)^{3/4}} = \frac{-(bc)^{3/4}}{(bc)^{3/4}} = -1 \text{ (mínimo)}$$

**Doolittle**  $[0, 1]$

$$A_{21}(j, k)^{(a,b,c,d)} = \Phi^2 = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{21}(j, k)^{(a=0,b,c,d=0)} = \frac{(0 * 0 - bc)^2}{(0 + b)((c + 0)(0 + c)(b + 0))} = \frac{(-bc)^2}{bccb} = \frac{(bc)^2}{(bc)^2} = 1 \text{ NO ES EL MÍNIMO, ES EL MÁXIMO}$$

**Eyraud** [-1, 0]

$$A_{22}(j, k)^{(a,b,c,d)} = \frac{a - (a + b)(a + c)}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{22}(j, k)^{(a=0,b,c,d=0)} = \frac{0 - (0+b)(0+c)}{(0+b)(c+0)(0+c)(b+0)} = \frac{-bc}{bccb} = \frac{-bc}{(bc)^2} = \frac{-1}{bc} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } bc = 1 \rightarrow A_{22}(j, k)^{(a=0,b,c,d=0)} = \frac{-1}{1} = -1 \text{ (mínimo)}$$

**Fager & McGowan** [-1, 1]

$$A_{23}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a + b)(a + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(a + b), (a + c)\}}}$$

$$A_{23}(j, k)^{(a=0,b,c,d=0)} = \frac{0}{\sqrt{(0 + b)(0 + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(0 + b), (0 + c)\}}} = -\frac{1}{2\sqrt{\text{máx.}\{(b), (c)\}}}$$

$$\text{Si } b > c \rightarrow A_{23}(j, k)^{(a=0,b,c,d=0)} = -\frac{1}{2\sqrt{b}} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b < c \rightarrow A_{23}(j, k)^{(a=0,b,c,d=0)} = -\frac{1}{2\sqrt{c}} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{23}(j, k)^{(a=0,b,c,d=0)} = -\frac{1}{2\sqrt{c}} \text{ NO ES EL MÍNIMO}$$

**Faith** [0, 1]

$$A_{24}(j, k)^{(a,b,c,d)} = \frac{a + \frac{d}{2}}{n} = \frac{a + \frac{d}{2}}{a + b + c + d}$$

$$A_{24}(j, k)^{(a=0,b,c,d=0)} = \frac{0 + \frac{0}{2}}{0 + b + c + 0} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Fleiss** [-∞, 1]

$$A_{25}(j, k)^{(a,b,c,d)} = \frac{(ad - bc)[(a + b)(b + d) + (a + c)(c + d)]}{2(a + b)(a + c)(b + d)(c + d)}$$

$$A_{25}(j, k)^{(a=0, b, c, d=0)} = \frac{(0 * 0 - bc)[(0 + b)(b + 0) + (0 + c)(c + 0)]}{2(0 + b)(0 + c)(b + 0)(c + 0)} = \frac{(-bc)[b^2 + c^2]}{2bc} \\ = \frac{-[b^2 + c^2]}{2bc} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{25}(j, k)^{(a=0, b, c, d=0)} = \frac{-[c^2 + c^2]}{2c^2} = \frac{-2c^2}{2c^2} = -1 \text{ NO ES EL MÍNIMO}$$

### Forbes\_1 [0, +∞]

$$A_{26}(j, k)^{(a, b, c, d)} = \frac{n * a}{(a + b)(a + c)} = \frac{(a + b + c + d)a}{(a + b)(a + c)}$$

$$A_{26}(j, k)^{(a=0, b, c, d=0)} = \frac{(0 + b + c + 0)0}{(0 + b)(0 + c)} = \frac{(b + c) * 0}{bc} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

### Forbes\_2 [-1, 1]

$$A_{27}(j, k)^{(a, b, c, d)} = \frac{n * a - (a + b)(a + c)}{n * \text{máx.}\{(a + b), (a + c)\} - (a + b)(a + c)} \\ = \frac{(a + b + c + d)a - (a + b)(a + c)}{(a + b + c + d) * \text{máx.}\{(a + b), (a + c)\} - (a + b)(a + c)}$$

$$A_{27}(j, k)^{(a=0, b, c, d=0)} = \frac{(0 + b + c + 0)0 - (0 + b)(0 + c)}{(0 + b + c + 0) * \text{máx.}\{(0 + b), (0 + c)\} - (0 + b)(0 + c)} \\ = \frac{(b + c) * 0 - (bc)}{(b + c) * \text{máx.}\{(b), (c)\} - bc} = \frac{-bc}{(b + c) * \text{máx.}\{(b), (c)\} - bc}$$

$$\text{Si } b > c \rightarrow A_{27}(j, k)^{(a=0, b, c, d=0)} = \frac{-bc}{(b + c)b - bc} = \frac{-c}{(b + c) - c} = \frac{-c}{b} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b < c \rightarrow A_{27}(j, k)^{(a=0, b, c, d=0)} = \frac{-bc}{(b + c)c - bc} = \frac{-b}{(b + c) - b} = \frac{-b}{c} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{27}(j, k)^{(a=0, b, c, d=0)} = \frac{-cc}{(c + c)c - cc} = \frac{-c^2}{2c^2 - c^2} = \frac{-c^2}{c^2} = -1 \text{ (mínimo)}$$

### Forbes\_3 [-1, +∞]

$$A_{28}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a + b)(a + c)}$$

$$A_{28}(j, k)^{(a=0, b, c, d=0)} = \frac{0 * 0 - bc}{(0 + b)(0 + c)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

### Fossum [0, +∞]

$$A_{29}(j, k)^{(a, b, c, d)} = \frac{n \left(a - \frac{1}{2}\right)^2}{(a + b)(a + c)} = \frac{(a + b + c + d) \left(a - \frac{1}{2}\right)^2}{(a + b)(a + c)}$$

$$A_{29}(j, k)^{(a=0, b, c, d=0)} = \frac{(0 + b + c + 0)(0 - \frac{1}{2})^2}{(0 + b)(0 + c)} = \frac{(b + c)\frac{1}{4}}{bc} = \frac{(b + c)}{4bc} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{29}(j, k)^{(a=0, b, c, d=0)} = \frac{c + c}{4cc} = \frac{2c}{4c^2} = \frac{1}{2c} < 1 \text{ NO ES EL MÍNIMO}$$

**Gilbert & Wells**  $[-\infty, +\infty]$

$$\begin{aligned} A_{30}(j, k)^{(a, b, c, d)} &= \log(a) - \log(n) - \log\left(\frac{a + b}{n}\right) - \log\left(\frac{a + c}{n}\right) \\ &= \log(a) - \log(a + b + c + d) - \log\left(\frac{a + b}{a + b + c + d}\right) - \log\left(\frac{a + c}{a + b + c + d}\right) \end{aligned}$$

$$\begin{aligned} A_{30}(j, k)^{(a=0, b, c, d=0)} &= \log(0) - \log(0 + b + c + 0) - \log\left(\frac{0 + b}{0 + b + c + 0}\right) - \log\left(\frac{0 + c}{0 + b + c + 0}\right) \\ &= \log(0) - \log(b + c) - \log\left(\frac{b}{b + c}\right) - \log\left(\frac{c}{b + c}\right) = \rightarrow -\infty \text{ (mínimo)} \end{aligned}$$

**Goodall**  $[0, 1]$

$$A_{31}(j, k)^{(a, b, c, d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a + d}{n}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a + d}{a + b + c + d}}$$

$$\begin{aligned} A_{31}(j, k)^{(a=0, b, c, d=0)} &= \frac{2}{\pi} \arcsin \sqrt{\frac{0 + 0}{0 + b + c + 0}} = \frac{2}{\pi} \arcsin \sqrt{\frac{0}{b + c}} = \frac{2}{\pi} \arcsin \sqrt{0} \\ &= 0 \text{ (mínimo)} \end{aligned}$$

**Goodman & Kruskal**  $[-1, 1]$

$$A_{32}(j, k)^{(a, b, c, d)} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

$$A_{32}(j, k)^{(a=0, b, c, d=0)} = \frac{2\text{mín.}\{(0), (0)\} - b - c}{2\text{mín.}\{(0), (0)\} + b + c} = \frac{2 * 0 - b - c}{2 * 0 + b + c} = \frac{-(b + c)}{b + c} = -1 \text{ (mínimo)}$$

**Hamann**  $[-1, 1]$

$$A_{33}(j, k)^{(a, b, c, d)} = \frac{a + d - b - c}{n} = \frac{a + d - b - c}{a + d + b + c}$$

$$A_{33}(j, k)^{(a=0, b, c, d=0)} = \frac{0 + 0 - b - c}{0 + 0 + b + c} = \frac{-(b + c)}{b + c} = -1 \text{ (mínimo)}$$

**Harris & Lahey**  $[0, n]$

$$A_{34}(j, k)^{(a,b,c,d)} = \frac{a(2d + b + c)}{2(a + b + c)} + \frac{d(2a + b + c)}{2(d + b + c)}$$

$$\begin{aligned} A_{34}(j, k)^{(a=0,b,c,d=0)} &= \frac{0(2 * 0 + b + c)}{2(0 + b + c)} + \frac{0(2 * 0 + b + c)}{2(0 + b + c)} = \frac{0(b + c)}{2(b + c)} + \frac{0(b + c)}{2(b + c)} \\ &= \frac{0}{2(b + c)} + \frac{0}{2(b + c)} = 0 + 0 = 0 \text{ (mínimo)} \end{aligned}$$

**Hawkins & Dotson** [0, 1]

$$A_{35}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{(a + b + c)} + \frac{d}{(d + b + c)} \right)$$

$$\begin{aligned} A_{35}(j, k)^{(a=0,b,c,d=0)} &= \frac{1}{2} \left( \frac{0}{(0 + b + c)} + \frac{0}{(0 + b + c)} \right) = \frac{1}{2} \left( \frac{0}{b + c} + \frac{0}{b + c} \right) = \frac{1}{2} (0 + 0) \\ &= \frac{1}{2} * 0 = 0 \text{ (mínimo)} \end{aligned}$$

**Hurlbert\_1** [-1, 1]

$$\begin{aligned} A_{36}(j, k)^{(a,b,c,d)} &= \pm \sqrt{\frac{\chi^2}{\chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2}} \\ &= \pm \sqrt{\frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2}} \end{aligned}$$

$$A_{36}(j, k)^{(a=0,b,c,d=0)} = \pm \sqrt{\frac{(0 + b + c + 0)(0 * 0 - bc)^2}{(0 + b)(c + 0)(0 + c)(b + 0) \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{(b + c)(bc)^2}{bccb \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{b + c}{\chi_{\text{máx.}}^2}}$$

$$\begin{aligned} \text{Si } \chi_{\text{máx.}}^2 = n = a + b + c + d = 0 + b + c + 0 = b + c \rightarrow A_{36}(j, k)^{(a=0,b,c,d=0)} &= \pm \sqrt{\frac{b + c}{b + c}} \\ &= -1 \text{ (mínimo) [con el signo de } (ad - bc)] \end{aligned}$$

**Hurlbert\_2** [-1, 1]

$$\begin{aligned} A_{37}(j, k)^{(a,b,c,d)} &= \pm \sqrt{\frac{\chi^2 - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \pm \sqrt{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2} - \chi_{\text{mín.}}^2} \\ &= \pm \sqrt{\frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2} - \chi_{\text{mín.}}^2} \end{aligned}$$

$$A_{37}(j, k)^{(a=0, b, c, d=0)} = \pm \sqrt{\frac{(0 + b + c + 0)(0 * 0 - bc)^2}{(0 + b)(c + 0)(0 + c)(b + 0)} - \chi_{\text{mín.}}^2}$$

$$= \pm \sqrt{\frac{(b + c)(bc)^2}{bccb} - \chi_{\text{mín.}}^2} = \pm \sqrt{\frac{(b + c) - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}}$$

Si  $\chi_{\text{máx.}}^2 = n = a + b + c + d = 0 + b + c + 0 = b + c$  y

$$\chi_{\text{mín.}}^2 = 0 \rightarrow A_{37}(j, k)^{(a=0, b, c, d=0)} = \pm \sqrt{\frac{(b + c) - 0}{(b + c) - 0}}$$

$$= -1 \text{ (mínimo) [con el signo de } (ad - bc)\text{]}$$

**Jaccard** [0, 1]

$$A_{38}(j, k)^{(a, b, c, d)} = \frac{a}{a + b + c}$$

$$A_{38}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Johnson\_1** [0, 2]

$$A_{39}(j, k)^{(a, b, c, d)} = \frac{a}{a + b} + \frac{a}{a + c}$$

$$A_{39}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{0 + b} + \frac{0}{0 + c} = \frac{0}{b} + \frac{0}{c} = 0 \text{ (mínimo)}$$

**Johnson\_2** [0, ∞]

$$A_{40}(j, k)^{(a, b, c, d)} = \frac{a}{2b}$$

$$A_{40}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{2b} = 0 \text{ (mínimo)}$$

**Kendall** [-1, 1]

$$A_{41}(j, k)^{(a, b, c, d)}$$

$$= \frac{2(ad - bc)}{\sqrt{(n^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)(n^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

$$= \frac{2(ad - bc)}{\sqrt{([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

$$\begin{aligned}
& A_{41}(j, k)^{(a=0, b, c, d=0)} \\
&= \frac{2(0 * 0 - bc)}{\sqrt{([0 + b + c + 0]^2 - 0^2 - b^2 - c^2 - 0^2 + 2b0 + 2c0)([0 + b + c + 0]^2 - 0^2 - b^2 - c^2 - 0^2 + 2c0 + 2b0)}} \\
&= \frac{-2bc}{-2bd} \\
&= \frac{\sqrt{([b + c]^2 - b^2 - c^2)([b + c]^2 - b^2 - c^2)}}{\sqrt{([b + c]^2 - b^2 - c^2)^2}} \\
&= \frac{-2bc}{[b + c]^2 - b^2 - c^2} = \frac{-2bc}{b^2 + 2bc + c^2 - b^2 - c^2} = \frac{-2bc}{2bc} = -1 \text{ (mínimo)}
\end{aligned}$$

**Kuder & Richardson**  $[-\infty, 1]$

$$\begin{aligned}
A_{42}(j, k)^{(a, b, c, d)} &= \frac{4(ad - bc)}{(a + b)(c + d) + (a + c)(b + d) + 2(ad - bc)} \\
A_{42}(j, k)^{(a=0, b, c, d=0)} &= \frac{4(0 * 0 - bc)}{(0 + b)(c + 0) + (0 + c)(b + 0) + 2(0 * 0 - bc)} \\
&= \frac{-4bc}{bc + cb + 2(-bc)} = \frac{-4bc}{2bc - 2bc} = \frac{-4bc}{0} \rightarrow -\infty \text{ (mínimo)}
\end{aligned}$$

**Kulczynski\_1**  $[0, +\infty]$

$$\begin{aligned}
A_{43}(j, k)^{(a, b, c, d)} &= \frac{a}{b + c} \\
A_{43}(j, k)^{(a=0, b, c, d=0)} &= \frac{0}{b + c} = 0 \text{ (mínimo)}
\end{aligned}$$

**Kulczynski\_2**  $[0, 1]$

$$\begin{aligned}
A_{44}(j, k)^{(a, b, c, d)} &= \frac{1}{2} \left( \frac{a}{a + b} + \frac{a}{a + c} \right) \\
A_{44}(j, k)^{(a=0, b, c, d=0)} &= \frac{1}{2} \left( \frac{0}{0 + b} + \frac{0}{0 + c} \right) = \frac{1}{2} \left( \frac{0}{b} + \frac{0}{c} \right) = \frac{1}{2} * 0 = 0 \text{ (mínimo)}
\end{aligned}$$

**Lamont & Grant**  $[0, 0.5]$

$$\begin{aligned}
A_{45}(j, k)^{(a, b, c, d)} &= \frac{a}{2a + b + c} \\
A_{45}(j, k)^{(a=0, b, c, d=0)} &= \frac{0}{2 * 0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}
\end{aligned}$$

**Legendre & Legendre**  $[0, 1]$

$$\begin{aligned}
A_{46}(j, k)^{(a, b, c, d)} &= \frac{3a}{3a + b + c} \\
A_{46}(j, k)^{(a=0, b, c, d=0)} &= \frac{3 * 0}{3 * 0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}
\end{aligned}$$

**Loevinger**  $[-\infty, 1]$

$$A_{47}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a+b)(b+d), (a+c)(c+d)\}}$$

$$A_{47}(j, k)^{(a=0,b,c,d=0)} = \frac{0 * 0 - bc}{\text{mín.}\{(0+b)(b+0), (0+c)(c+0)\}} = \frac{-bc}{\text{mín.}\{(bb), (cc)\}}$$

$$\text{Si } b < c \rightarrow A_{47}(j, k)^{(a=0,b,c,d=0)} = \frac{-bc}{bb} = \frac{-c}{b} \text{ NO ES ES MÍNIMO}$$

$$\text{Si } b > c \rightarrow A_{47}(j, k)^{(a=0,b,c,d=0)} = \frac{-bc}{cc} = \frac{-b}{c} \text{ NO ES ES MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{47}(j, k)^{(a=0,b,c,d=0)} = \frac{-cc}{cc} = -1 \text{ NO ES ES MÍNIMO}$$

**Maxwell & Pilliner**  $[-1, 1]$

$$A_{48}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(c+d) + (a+c)(b+d)}$$

$$A_{48}(j, k)^{(a=0,b,c,d=0)} = \frac{2(0 * 0 - bc)}{(0+b)(c+0) + (0+c)(b+0)} = \frac{-2bc}{bc + cb} = \frac{-2bc}{2bc} = -1 \text{ (mínimo)}$$

**McConnaughey**  $[-1, 1]$

$$A_{49}(j, k)^{(a,b,c,d)} = \frac{a^2 - bc}{(a+b)(a+c)}$$

$$A_{49}(j, k)^{(a=0,b,c,d=0)} = \frac{0^2 - bc}{(0+b)(0+c)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Michael**  $[-1, 1]$

$$A_{50}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a+d)^2 + (b+c)^2}$$

$$A_{50}(j, k)^{(a=0,b,c,d=0)} = \frac{4(0 * 0 - bc)}{(0+0)^2 + (b+c)^2} = \frac{-4bc}{(b+c)^2} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b = c \rightarrow A_{50}(j, k)^{(a=0,b,c,d=0)} = \frac{-4cc}{(c+c)^2} = \frac{-4c^2}{(2c)^2} = \frac{-4c^2}{4c^2} = -1 \text{ (mínimo)}$$

**Mountford**  $[0, 2]$

$$A_{51}(j, k)^{(a,b,c,d)} = \frac{a}{\frac{1}{2}(ab + ac) + bc}$$

$$A_{51}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{\frac{1}{2}(0b + 0c) + bc} = \frac{0}{\frac{1}{2} * 0 + bc} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Ochiai** [0, 1]

$$A_{52}(j, k)^{(a, b, c, d)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

$$A_{52}(j, k)^{(a=0, b, c, d=0)} = \frac{0}{\sqrt{(0+b)(0+c)}} = \frac{0}{\sqrt{bc}} = 0 \text{ (mínimo)}$$

**Pearson & Heron** [-1, 1]

$$A_{53}(j, k)^{(a, b, c, d)} = \cos \frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{53}(j, k)^{(a=0, b, c, d=0)} = \cos \frac{180\sqrt{bc}}{\sqrt{00} + \sqrt{bc}} = \cos \frac{180\sqrt{bc}}{\sqrt{bc}} = \cos(180) = -1 \text{ (mínimo)}$$

**Pearson\_1** [0, n]

$$A_{54}(j, k)^{(a, b, c, d)} = \chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\begin{aligned} A_{54}(j, k)^{(a=0, b, c, d=0)} &= \frac{(0+b+c+0)(0*0 - bc)^2}{(0+b)(c+0)(0+c)(b+0)} = \frac{(b+c)(-bc)^2}{bccb} = \frac{(b+c)(bc)^2}{(bc)^2} \\ &= b+c = n = a+b+c+d \\ &= 0+b+c+0 \text{ NO ES EL MÍNIMO, ES EL MÁXIMO} \end{aligned}$$

**Pearson\_2** [0, 1]

$$A_{55}(j, k)^{(a, b, c, d)} = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{\frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{(a+b+c+d) + \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

$$\begin{aligned}
A_{55}(j, k)^{(a=0, b, c, d=0)} &= \sqrt{\frac{\frac{(0+b+c+0)(0*0-bc)^2}{(0+b)(c+0)(0+c)(b+0)}}{(0+b+c+0) + \frac{(0+b+c+0)(00-bc)^2}{(0+b)(c+0)(0+c)(b+0)}}} \\
&= \sqrt{\frac{\frac{(b+c)(bc)^2}{bccb}}{(b+c) + \frac{(b+c)(bc)^2}{bccb}}} = \sqrt{\frac{\frac{(b+c)(bc)^2}{(bc)^2}}{(b+c) + \frac{(b+c)(bc)^2}{(bc)^2}}} \sqrt{\frac{(b+c)}{(b+c) + (b+c)}} \\
&= \sqrt{\frac{(b+c)}{2(b+c)}} = \sqrt{\frac{1}{2}} = 0,7071 \text{ NO ES EL MÍNIMO}
\end{aligned}$$

**Peirce\_1** [0, 1]

$$A_{56}(j, k)^{(a, b, c, d)} = \frac{ab + bc}{ab + 2bc + cd}$$

$$A_{56}(j, k)^{(a=0, b, c, d=0)} = \frac{0b + bc}{0b + 2bc + c0} = \frac{bc}{2bc} = \frac{1}{2} = 0,5 \text{ NO ES EL MÍNIMO}$$

**Peirce\_2** [-1, 1]

$$A_{57}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a+c)(b+d)}$$

$$A_{57}(j, k)^{(a=0, b, c, d=0)} = \frac{0*0 - bc}{(0+c)(b+0)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Peirce\_3** [-1, 1]

$$A_{58}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a+b)(c+d)}$$

$$A_{58}(j, k)^{(a=0, b, c, d=0)} = \frac{0*0 - bc}{(0+b)(c+0)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Roger & Tanimoto** [0, 1]

$$A_{59}(j, k)^{(a, b, c, d)} = \frac{a+d}{a+2(b+c)+d}$$

$$A_{59}(j, k)^{(a=0, b, c, d=0)} = \frac{0+0}{0+2(b+c)+0} = \frac{0}{2(b+c)} = 0 \text{ (mínimo)}$$

**Rogot & Goldberg** [0, 1]

$$A_{60}(j, k)^{(a,b,c,d)} = \frac{a}{2a + b + c} + \frac{d}{2d + b + c}$$

$$A_{60}(j, k)^{(a=0,b,c,d=0)} = \frac{0}{2 * 0 + b + c} + \frac{0}{2 * 0 + b + c} = \frac{0}{b + c} + \frac{0}{b + c} = 0 + 0 = 0 \text{ (mínimo)}$$

**Russell & Rao** [0, 1]

$$A_{61}(j, k)^{(a,b,c,d)} = \frac{a}{n} = \frac{a}{a + b + c + d}$$

$$A_{61}(j, k)^{(a=0,b,c,d=0)} = \frac{0}{0 + b + c + 0} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Scott** [-1, 1]

$$A_{62}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(2d + b + c)}$$

$$\begin{aligned} A_{62}(j, k)^{(a=0,b,c,d=0)} &= \frac{4(0 * 0 - bc) - (b - c)^2}{(2 * 0 + b + c)(2 * 0 + b + c)} = \frac{-4bc - (b - c)^2}{(b + c)^2} \\ &= \frac{-4bc - (b^2 - 2bc + c^2)}{b^2 + 2bc + c^2} = \frac{-b^2 - c^2 - 2bc}{b^2 + c^2 + 2bc} = \frac{-(b^2 + c^2 + 2bc)}{(b^2 + c^2 + 2bc)} \\ &= -1 \text{ (mínimo)} \end{aligned}$$

**Simpson** [0, 1]

$$A_{63}(j, k)^{(a,b,c,d)} = \frac{a}{\text{mín.}\{(a + b), (a + c)\}}$$

$$A_{63}(j, k)^{(a=0,b,c,d=0)} = \frac{0}{\text{mín.}\{(0 + b), (0 + c)\}} = \frac{0}{\text{mín.}\{(b), (c)\}} = 0 \text{ (mínimo)}$$

**Sokal & Michener** [0, 1]

$$A_{64}(j, k)^{(a,b,c,d)} = \frac{a + d}{n} = \frac{a + d}{a + b + c + d}$$

$$A_{64}(j, k)^{(a=0,b,c,d=0)} = \frac{0 + 0}{0 + b + c + 0} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_1** [0, 1]

$$A_{65}(j, k)^{(a,b,c,d)} = \frac{a}{a + 2(b + c)}$$

$$A_{65}(j, k)^{(a=0,b,c,d=0)} = \frac{0}{0 + 2(b + c)} = \frac{0}{2(b + c)} = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_2** [0, 1]

$$A_{66}(j, k)^{(a,b,c,d)} = \frac{1}{4} \left( \frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{b+d} \right)$$

$$A_{66}(j, k)^{(a=0,b,c,d=0)} = \frac{1}{4} \left( \frac{0}{0+b} + \frac{0}{0+c} + \frac{0}{c+0} + \frac{0}{b+0} \right) = \frac{1}{4} \left( \frac{0}{b} + \frac{0}{c} + \frac{0}{c} + \frac{0}{b} \right) = \frac{1}{4} * 0 = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_3** [0, +∞]

$$A_{67}(j, k)^{(a,b,c,d)} = \frac{a+d}{b+c}$$

$$A_{67}(j, k)^{(a=0,b,c,d=0)} = \frac{0+0}{b+c} = \frac{0}{b+c} = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_4** [0, 1]

$$A_{68}(j, k)^{(a,b,c,d)} = \frac{a+d}{a + \frac{1}{2}(b+c) + d}$$

$$A_{68}(j, k)^{(a=0,b,c,d=0)} = \frac{0+0}{0 + \frac{1}{2}(b+c) + 0} = \frac{0}{\frac{1}{2}(b+c)} = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_5** [0, 1]

$$A_{69}(j, k)^{(a,b,c,d)} = \frac{ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$A_{69}(j, k)^{(a=0,b,c,d=0)} = \frac{0*0}{\sqrt{(0+b)(c+0)(0+c)(b+0)}} = \frac{0}{\sqrt{bccb}} = \frac{0}{\sqrt{(bc)^2}} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Sorensen** [0, 1]

$$A_{70}(j, k)^{(a,b,c,d)} = \frac{2a}{2a+b+c}$$

$$A_{70}(j, k)^{(a=0,b,c,d=0)} = \frac{2*0}{2*0+b+c} = \frac{0}{b+c} = 0 \text{ (mínimo)}$$

**Sorgenfrei** [0, 1]

$$A_{71}(j, k)^{(a,b,c,d)} = \frac{a^2}{(a+b)(a+c)}$$

$$A_{71}(j, k)^{(a=0, b, c, d=0)} = \frac{0^2}{(0+b)(0+c)} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Stiles**  $[-\infty, +\infty]$

$$\begin{aligned} A_{72}(j, k)^{(a, b, c, d)} &= \log \frac{n(|ad - bc| - \frac{n}{2})^2}{(a+b)(a+c)(b+d)(c+d)} \\ &= \log \frac{(a+b+c+d)(|ad - bc| - \frac{a+b+c+d}{2})^2}{(a+b)(a+c)(b+d)(c+d)} \end{aligned}$$

$$\begin{aligned} A_{72}(j, k)^{(a=0, b, c, d=0)} &= \log \frac{(0+b+c+0) \left( |0*0 - bc| - \frac{0+b+c+0}{2} \right)^2}{(0+b)(0+c)(b+0)(c+0)} \\ &= \log \frac{(b+c)(bc - \frac{b+c}{2})^2}{bcbc} = \log \frac{(b+c) \left( \frac{2bc - (b+c)}{2} \right)^2}{(bc)^2} \\ &= \log \frac{(b+c)(2bc - (b+c))^2}{(2bc)^2} \text{ NO ES EL MÍNIMO} \end{aligned}$$

$$\begin{aligned} \text{Si } b = c \rightarrow A_{72}(j, k)^{(a=0, b, c, d=0)} &= \log \frac{(c+c)(2cc - (c+c))^2}{(2cc)^2} = \log \frac{2c(2c^2 - 2c)^2}{(2c^2)^2} \\ &= \log \frac{(2c)[2c(c-1)]^2}{4c^4} = \log \frac{(2c)(2c)^2(c-1)^2}{4c^4} = \log \frac{(2c)(4c^2)(c-1)^2}{4c^4} \\ &= \log \frac{2(c-1)^2}{c} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Tarwid**  $[-1, 1]$

$$A_{73}(j, k)^{(a, b, c, d)} = \frac{na - (a+b)(a+c)}{na + (a+b)(a+c)} = \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d)a + (a+b)(a+c)}$$

$$\begin{aligned} A_{73}(j, k)^{(a=0, b, c, d=0)} &= \frac{(0+b+c+0)0 - (0+b)(0+c)}{(0+b+c+0)0 + (0+b)(0+c)} = \frac{0 - bc}{0 + bc} = \frac{-bc}{bc} \\ &= -1 \text{ (mínimo)} \end{aligned}$$

**Tschuproff**  $[0, 1]$

$$\begin{aligned} A_{74}(j, k)^{(a, b, c, d)} &= \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{\frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{a+b+c+d}} \\ &= \sqrt{\frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}} \end{aligned}$$

$$\begin{aligned}
A_{74}(j, k)^{(a=0, b, c, d=0)} &= \sqrt{\frac{(0+b+c+0)(0*0-bc)^2}{(0+b)(c+0)(0+c)(b+0)}} = \sqrt{\frac{(b+c)(-bc)^2}{bccb}} = \sqrt{\frac{(b+c)(bc)^2}{(bc)^2}} \\
&= \sqrt{\frac{b+c}{b+c}} = \sqrt{1} = 1 \text{ NO ES EL MÍNIMO, ES EL MÁXIMO}
\end{aligned}$$

**Yule \_1** [-1, 1]

$$A_{75}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \phi$$

$$\begin{aligned}
A_{75}(j, k)^{(a=0, b, c, d=0)} &= \frac{0*0 - bc}{\sqrt{(0+b)(c+0)(0+c)(b+0)}} = \frac{-bc}{\sqrt{bccb}} = \frac{-bc}{\sqrt{b^2c^2}} = \frac{-bc}{bc} \\
&= -1 \text{ (mínimo)}
\end{aligned}$$

**Yule \_2** [-1, 1]

$$A_{76}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{ad + bc}$$

$$A_{76}(j, k)^{(a=0, b, c, d=0)} = \frac{0*0 - bc}{0*0 + bc} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Yule \_3** [-1, 1]

$$A_{77}(j, k)^{(a, b, c, d)} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{77}(j, k)^{(a=0, b, c, d=0)} = \frac{\sqrt{0*0} - \sqrt{bc}}{\sqrt{0*0} + \sqrt{bc}} = \frac{-\sqrt{bc}}{\sqrt{bc}} = -1 \text{ (mínimo)}$$



## ANEXO 2

Propiedad 3b: Existencia de valores máximos

$$\text{máx } A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)}$$

**Anónimo\_1** [-1, 1]

$$A_1(j, k)^{(a,b,c,d)} = \frac{2a - (b + c)}{2a + (b + c)}$$

$$A_1(j, k)^{(a,b=c=0,d)} = \frac{2a - (0 + 0)}{2a + (0 + 0)} = \frac{2a}{2a} = 1 \text{ (máximo)}$$

**Anónimo\_2** [0, 1]

$$A_2(j, k)^{(a,b,c,d)} = \frac{2d}{2d + b + c}$$

$$A_2(j, k)^{(a,b=c=0,d)} = \frac{2d}{2d + 0 + 0} = \frac{2d}{2d} = 1 \text{ (máximo)}$$

**Anónimo\_3** [-1, 1]

$$A_3(j, k)^{(a,b,c,d)} = \frac{2d - (b + c)}{2d + (b + c)}$$

$$A_3(j, k)^{(a,b=c=0,d)} = \frac{2d - (0 + 0)}{2d + (0 + 0)} = \frac{2d}{2d} = 1 \text{ (máximo)}$$

**Anónimo\_4** [0, 1]

$$A_4(j, k)^{(a,b,c,d)} = \frac{4ad}{4ad + (a + d)(b + c)}$$

$$A_4(j, k)^{(a,b=c=0,d)} = \frac{4ad}{4ad + (a + d)(0 + 0)} = \frac{4ad}{4ad + 0} = 1 \text{ (máximo)}$$

**Anónimo\_5** [-1, +∞]

$$A_5(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a + b)(a + c), (c + d)(b + d)\}}$$

$$\begin{aligned} A_5(j, k)^{(a,b=c=0,d)} &= \frac{ad - 0 * 0}{\text{mín}\{(a + 0)(a + 0), (0 + d)(0 + d)\}} = \frac{ad}{\text{mín}\{(aa), (dd)\}} \\ &= \frac{ad}{\text{mín}\{(a^2), (d^2)\}} \end{aligned}$$

$$\text{Si } a < d \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \frac{ad}{a^2} = \frac{d}{a} \leq \infty \text{ NO ES EL MÁXIMO}$$

$$\text{Si } a > d \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \frac{ad}{d^2} = \frac{a}{d} \leq \infty \text{ NO ES EL MÁXIMO}$$

$$\text{Si } a = d \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \frac{d^2}{d^2} = 1 \text{ NO ES EL MÁXIMO}$$

### Baroni-Urbani & Buser\_1 [0, 1]

$$A_6(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

$$A_6(j, k)^{(a,b=c=0,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + 0 + 0} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a} = 1 \text{ (máximo)}$$

### Baroni-Urbani & Buser\_2 [-1, 1]

$$A_7(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a - b - c}{\sqrt{ad} + a + b + c}$$

$$A_7(j, k)^{(a,b=c=0,d)} = \frac{\sqrt{ad} + a - 0 - 0}{\sqrt{ad} + a + 0 + 0} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a} = 1 \text{ (máximo)}$$

### Braun-Blanquet [0, 1]

$$A_8(j, k)^{(a,b,c,d)} = \frac{a}{\text{máx.}\{(a+b), (a+c)\}}$$

$$A_8(j, k)^{(a,b=c=0,d)} = \frac{a}{\text{máx.}\{(a+0), (a+0)\}} = \frac{a}{\text{máx.}\{(a), (a)\}} = \frac{a}{a} = 1 \text{ (máximo)}$$

### Clement [0, n]

$$A_9(j, k)^{(a,b,c,d)} = \frac{a(c+d)}{(a+b)} + \frac{d(a+b)}{(c+d)}$$

$$A_9(j, k)^{(a,b=c=0,d)} = \frac{a(0+d)}{(a+0)} + \frac{d(a+0)}{(0+d)} = \frac{ad}{a} + \frac{da}{d} = d + a$$

$$n = a + b + c + d = a + 0 + 0 + d = a + d$$

$$A_9(j, k)^{(a,b=c=0,d)} = d + a = n \text{ (máximo)}$$

### Clifford & Stephenson [0, 1]

$$A_{10}(j, k)^{(a,b,c,d)} = \frac{\chi^2}{n} = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$A_{10}(j, k)^{(a,b=c=0,d)} = \frac{(ad - 00)^2}{(a+0)(0+d)(a+0)(0+d)} = \frac{(ad)^2}{adad} = \frac{(ad)^2}{(ad)^2} = 1 \text{ (máximo)}$$

**Cohen** [-1, 1]

$$A_{11}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(b+d) + (a+c)(c+d)}$$

$$A_{11}(j, k)^{(a,b=c=0,d)} = \frac{2(ad - 00)}{(a+0)(0+d) + (a+0)(0+d)} = \frac{2ad}{ad + ad} = \frac{2ad}{2ad} = 1 \text{ (máximo)}$$

**Cole\_1** [-1, 1]

$$A_{12}(j, k)^{(a,b,c,d)} = \frac{\sqrt{2}(ad - bc)}{\sqrt{(ad - bc)^2 + (a+b)(c+d)(a+c)(b+d)}}$$

$$\begin{aligned} A_{12}(j, k)^{(a,b=c=0,d)} &= \frac{\sqrt{2}(ad - 00)}{\sqrt{(ad - 00)^2 + (a+0)(0+d)(a+0)(0+d)}} = \frac{\sqrt{2}ad}{\sqrt{(ad)^2 + (ad)^2}} \\ &= \frac{\sqrt{2}ad}{\sqrt{2(ad)^2}} = \frac{\sqrt{2}ad}{\sqrt{2}ad} = 1 \text{ (máximo)} \end{aligned}$$

**Cole\_2** [0, 1]

$$A_{13}(j, k)^{(a,b,c,d)} = \frac{\phi}{\phi_{\text{máx}}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \cdot \frac{\phi_{\text{máx}}}{\phi_{\text{máx}}}$$

$$A_{13}(j, k)^{(a,b=c=0,d)} = \frac{ad - 00}{\sqrt{(a+0)(0+d)(a+0)(0+d)}} \cdot \frac{\phi_{\text{máx}}}{\phi_{\text{máx}}} = \frac{ad}{\sqrt{(ad)^2}} \cdot \frac{\phi_{\text{máx}}}{\phi_{\text{máx}}} = \frac{1}{1} = 1$$

Si  $\phi_{\text{máx}} = 1 \rightarrow A_{13}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$

**Cole\_3**  $[-\infty, 1]$

$$A_{14}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+b)(b+d)}$$

$$A_{14}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{(a+0)(0+d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

**Cole\_4**  $[-\infty, 1]$

$$A_{15}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + c)(c + d)}$$

$$A_{15}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{(a + 0)(0 + d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

**Dennis**  $[-\infty, +\infty]$

$$A_{16}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a + b + c + d)(a + b)(a + c)}}$$

$$A_{16}(j, k)^{(a,b=c=0,d)} = \frac{ad - 00}{\sqrt{(a + 0 + 0 + d)(a + 0)(a + 0)}} = \frac{ad}{\sqrt{(a + d)a^2}} = \frac{ad}{a * \sqrt{(a + d)}}$$

$$= \frac{d}{\sqrt{a + d}} \text{ NO ES EL MÁXIMO}$$

**Dice\_1**  $[0, 1]$

$$A_{17}(j, k)^{(a,b,c,d)} = \frac{a}{a + b}$$

$$A_{17}(j, k)^{(a,b=c=0,d)} = \frac{a}{a + 0} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Dice\_2**  $[0, 1]$

$$A_{18}(j, k)^{(a,b,c,d)} = \frac{a}{a + c}$$

$$A_{18}(j, k)^{(a,b=c=0,d)} = \frac{a}{a + 0} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Dice\_3**  $[0, 2]$

$$A_{19}(j, k)^{(a,b,c,d)} = \frac{2a}{(a + b)(a + c)}$$

$$A_{19}(j, k)^{(a,b=c=0,d)} = \frac{2a}{(a + 0)(a + 0)} = \frac{2a}{a^2} = \frac{2}{a}$$

$$\text{Si } a = 1 \rightarrow A_5(j, k)^{(a,b=c=0,d)} = 2 \text{ (máximo)}$$

$$\text{Si } a \neq 1 \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \frac{2}{a} < 2 \text{ NO ES EL MÁXIMO}$$

**Digby**  $[-1, 1]$

$$A_{20}(j, k)^{(a,b,c,d)} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}}$$

$$A_{20}(j, k)^{(a,b=c=0,d)} = \frac{(ad)^{\frac{3}{4}} - (0 * 0)^{\frac{3}{4}}}{(ad)^{\frac{3}{4}} + (0 * 0)^{\frac{3}{4}}} = \frac{(ad)^{\frac{3}{4}}}{(ad)^{\frac{3}{4}}} = 1 \text{ (máximo)}$$

**Doolittle** [0, 1]

$$A_{21}(j, k)^{(a,b,c,d)} = \Phi^2 = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{21}(j, k)^{(a,b=c=0,d)} = \frac{(ad - 00)^2}{(a + 0)((0 + d)(a + 0)(0 + d))} = \frac{(ad)^2}{adad} = \frac{(ad)^2}{(ad)^2} = 1 \text{ (máximo)}$$

**Eyraud** [-1, 0]

$$A_{22}(j, k)^{(a,b,c,d)} = \frac{a - (a + b)(a + c)}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{22}(j, k)^{(a,b=c=0,d)} = \frac{a - (a+0)(a+0)}{(a+0)(0+d)(a+0)(0+d)} = \frac{a - aa}{adad} = \frac{a(1-a)}{a^2d^2} = \frac{1-a}{ad^2} \text{ NO ES EL MÁXIMO}$$

**Fager & McGowan** [-1, 1]

$$A_{23}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a + b)(a + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(a + b), (a + c)\}}}$$

$$A_{23}(j, k)^{(a,b=c=0,d)} = \frac{a}{\sqrt{(a + 0)(a + 0)}} - \frac{1}{2\sqrt{\text{máx.}\{(a + 0), (a + 0)\}}} \\ = \frac{a}{\sqrt{a^2}} - \frac{1}{2\sqrt{\text{máx.}\{(a), (a)\}}} = \frac{a}{a} - \frac{1}{2\sqrt{a}} = 1 - \frac{1}{2\sqrt{a}} \text{ NO ES EL MÁXIMO}$$

**Faith** [0, 1]

$$A_{24}(j, k)^{(a,b,c,d)} = \frac{a + \frac{d}{2}}{n} = \frac{a + \frac{d}{2}}{a + b + c + d}$$

$$A_{24}(j, k)^{(a,b=c=0,d)} = \frac{a + \frac{d}{2}}{a + 0 + 0 + d} = \frac{a + \frac{d}{2}}{a + d} = \frac{2a + d}{2(a + d)} = \frac{2a + d}{2} \cdot \frac{1}{a + d} = \frac{2a + d}{2a + 2d} \\ < 1 \text{ NO ES EL MÁXIMO}$$

$$\text{Si } a = d \rightarrow A_{24}(j, k)^{(a,b=c=0,d)} = \frac{2d + d}{2d + 2d} = \frac{3d}{4d} = \frac{3}{4} < 1 \text{ NO ES EL MÁXIMO}$$

**Fleiss**  $[-\infty, 1]$

$$A_{25}(j, k)^{(a,b,c,d)} = \frac{(ad - bc)[(a + b)(b + d) + (a + c)(c + d)]}{2(a + b)(a + c)(b + d)(c + d)}$$

$$A_{25}(j, k)^{(a,b=c=0,d)} = \frac{(ad - 00)[(a+0)(0+d) + (a+0)(0+d)]}{2(a+0)(a+0)(0+d)(0+d)} = \frac{(ad)[ad + ad]}{2aadd}$$

$$= \frac{(ad)(2ad)}{2a^2d^2} = \frac{2a^2d^2}{2a^2d^2} = 1 \quad (\text{máximo})$$

**Forbes\_1** [0, +∞]

$$A_{26}(j, k)^{(a,b,c,d)} = \frac{n * a}{(a+b)(a+c)} = \frac{(a+b+c+d)a}{(a+b)(a+c)}$$

$$A_{26}(j, k)^{(a,b=c=0,d)} = \frac{(a+0+0+d)a}{(a+0)(a+0)} = \frac{(a+d)a}{aa} = \frac{a+d}{a} \quad \text{NO ES EL MÁXIMO}$$

**Forbes\_2** [-1, 1]

$$A_{27}(j, k)^{(a,b,c,d)} = \frac{n * a - (a+b)(a+c)}{n * \text{máx.}\{(a+b), (a+c)\} - (a+b)(a+c)}$$

$$= \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d) * \text{máx.}\{(a+b), (a+c)\} - (a+b)(a+c)}$$

$$A_{27}(j, k)^{(a,b=c=0,d)} = \frac{(a+0+0+d)a - (a+0)(a+0)}{(a+0+0+d) * \text{máx.}\{(a+0), (a+0)\} - (a+0)(a+0)}$$

$$= \frac{(a+d)a - aa}{(a+d) * \text{máx.}\{(a), (a)\} - aa} = \frac{(a+d)a - a^2}{(a+d)a - a^2} = 1 \quad (\text{máximo})$$

**Forbes\_3** [-1, +∞]

$$A_{28}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+b)(a+c)}$$

$$A_{28}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{(a+0)(a+0)} = \frac{ad}{aa} = \frac{d}{a} \quad \text{NO ES EL MÁXIMO}$$

$$\text{Si } a = d \rightarrow A_{28}(j, k)^{(a,b=c=0,d)} = \frac{d}{d} = 1 \quad \text{NO ES EL MÁXIMO}$$

**Fossum** [0, +∞]

$$A_{29}(j, k)^{(a,b,c,d)} = \frac{n \left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)} = \frac{(a+b+c+d) \left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)}$$

$$A_{29}(j, k)^{(a,b=c=0,d)} = \frac{(a+0+0+d) \left(a - \frac{1}{2}\right)^2}{(a+0)(a+0)} = \frac{(a+d) \left(\frac{2a-1}{2}\right)^2}{aa}$$

$$= \frac{(a+d)(2a-1)^2}{4a^2} \quad \text{NO ES EL MÁXIMO}$$

**Gilbert & Wells**  $[-\infty, +\infty]$ 

$$A_{30}(j, k)^{(a,b,c,d)} = \log(a) - \log(n) - \log\left(\frac{a+b}{n}\right) - \log\left(\frac{a+c}{n}\right)$$

$$= \log(a) - \log(a+b+c+d) - \log\left(\frac{a+b}{a+b+c+d}\right) - \log\left(\frac{a+c}{a+b+c+d}\right)$$

$$A_{30}(j, k)^{(a,b=c=0,d)}$$

$$= \log(a) - \log(a+0+0+d) - \log\left(\frac{a+0}{a+0+0+d}\right) - \log\left(\frac{a+0}{a+0+0+d}\right)$$

$$= \log(a) - \log(a+d) - \log\left(\frac{a}{a+d}\right)$$

$$- \log\left(\frac{a}{a+d}\right) = \log\left(\frac{a}{a+d}\right) - 2\log\left(\frac{a}{a+d}\right)$$

$$= -\log\left(\frac{a}{a+d}\right) \text{ NO ES EL MÁXIMO}$$

**Goodall**  $[0, 1]$ 

$$A_{31}(j, k)^{(a,b,c,d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{n}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+c+d}}$$

$$A_{31}(j, k)^{(a,b=c=0,d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+0+0+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+d}} = \frac{2}{\pi} \arcsin \sqrt{1} = \frac{2}{\pi} \frac{\pi}{2} =$$

$$= 1 \text{ (máximo)}$$

**Goodman & Kruskal**  $[-1, 1]$ 

$$A_{32}(j, k)^{(a,b,c,d)} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

$$A_{32}(j, k)^{(a,b=c=0,d)} = \frac{2\text{mín.}\{(a), (d)\} - 0 - 0}{2\text{mín.}\{(a), (d)\} + 0 + 0} = \frac{2\text{mín.}\{(a), (d)\}}{2\text{mín.}\{(a), (d)\}} = 1 \text{ (máximo)}$$

**Hamann**  $[-1, 1]$ 

$$A_{33}(j, k)^{(a,b,c,d)} = \frac{a+d-b-c}{n} = \frac{a+d-b-c}{a+d+b+c}$$

$$A_{33}(j, k)^{(a,b=c=0,d)} = \frac{a+d-0-0}{a+d+0+0} = \frac{a+d}{a+d} = 1 \text{ (máximo)}$$

**Harris & Lahey**  $[0, n]$ 

$$A_{34}(j, k)^{(a,b,c,d)} = \frac{a(2d+b+c)}{2(a+b+c)} + \frac{d(2a+b+c)}{2(d+b+c)}$$

$$A_{34}(j, k)^{(a,b=c=0,d)} = \frac{a(2d+0+0)}{2(a+0+0)} + \frac{d(2a+0+0)}{2(d+0+0)} = \frac{a(2d)}{2(a)} + \frac{d(2a)}{2(d)} = \frac{2ad}{2a} + \frac{2ad}{2d} = d + a$$

$$n = a + b + c + d = a + 0 + 0 + d = a + d$$

$$A_{34}(j, k)^{(a,b=c=0,d)} = d + a = n \text{ (máximo)}$$

### Hawkins & Dotson [0, 1]

$$A_{35}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{(a+b+c)} + \frac{d}{(d+b+c)} \right)$$

$$A_{35}(j, k)^{(a,b=c=0,d)} = \frac{1}{2} \left( \frac{a}{(a+0+0)} + \frac{d}{(d+0+0)} \right) = \frac{1}{2} \left( \frac{a}{a} + \frac{d}{d} \right) = \frac{1}{2} (1 + 1) = \frac{1}{2} * 2 = 1 \text{ (máximo)}$$

### Hurlbert\_1 [-1, 1]

$$A_{36}(j, k)^{(a,b,c,d)} = \pm \sqrt{\frac{\chi^2}{\chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{\text{máx.}}^2}}$$

$$= \pm \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{\text{máx.}}^2}}$$

$$A_{36}(j, k)^{(a,b=c=0,d)} = \pm \sqrt{\frac{(a+0+0+d)(ad-0*0)^2}{(a+0)(0+d)(a+0)(0+d) \chi_{\text{máx.}}^2}}$$

$$= \pm \sqrt{\frac{(a+d)(ad)^2}{adad \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{(a+d)(ad)^2}{(ad)^2 \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{a+d}{\chi_{\text{máx.}}^2}}$$

$$\text{Si } \chi_{\text{máx.}}^2 = n = a + b + c + d = a + 0 + 0 + d = a + d \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \pm \sqrt{\frac{a+d}{a+d}}$$

$$= 1 \text{ (máximo)}$$

### Hurlbert\_2 [-1, 1]

$$A_{37}(j, k)^{(a,b,c,d)} = \pm \sqrt{\frac{\chi^2 - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \pm \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) - \chi_{\text{mín.}}^2}}$$

$$= \pm \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) - \chi_{\text{mín.}}^2}}$$

$$A_5(j, k)^{(a,b=c=0,d)} = \pm \sqrt{\frac{(a+0+0+d)(ad-0*0)^2}{(a+0)(0+d)(a+0)(0+d)} - \chi_{\min}^2} \\ = \pm \sqrt{\frac{(a+d)(ad)^2}{adad} - \chi_{\min}^2} = \pm \sqrt{\frac{(a+d)(ad)^2}{(ad)^2} - \chi_{\min}^2} = \pm \sqrt{\frac{(a+d) - \chi_{\min}^2}{\chi_{\max}^2 - \chi_{\min}^2}}$$

Si  $\chi_{\max}^2 = n = a + b + c + d = a + 0 + 0 + d = a + d$  y

$$\chi_{\min}^2 = 0 \rightarrow A_5(j, k)^{(a,b=c=0,d)} = \pm \sqrt{\frac{(a+d) - 0}{(a+d) - 0}} = 1 \text{ (máximo)}$$

**Jaccard** [0, 1]

$$A_{38}(j, k)^{(a,b,c,d)} = \frac{a}{a+b+c}$$

$$A_{38}(j, k)^{(a,b=c=0,d)} = \frac{a}{a+0+0} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Johnson\_1** [0, 2]

$$A_{39}(j, k)^{(a,b,c,d)} = \frac{a}{a+b} + \frac{a}{a+c}$$

$$A_{39}(j, k)^{(a,b=c=0,d)} = \frac{a}{a+0} + \frac{a}{a+0} = \frac{a}{a} + \frac{a}{a} = 1 + 1 = 2 \text{ (máximo)}$$

**Johnson\_2** [0, +∞]

$$A_{40}(j, k)^{(a,b,c,d)} = \frac{a}{2b}$$

$$A_{40}(j, k)^{(a,b=c=0,d)} = \frac{a}{2*0} = \frac{a}{0} \rightarrow +\infty \text{ (máximo)}$$

**Kendall** [-1, 1]

$$A_{41}(j, k)^{(a,b,c,d)}$$

$$= \frac{2(ad - bc)}{\sqrt{(n^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)(n^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}} \\ = \frac{2(ad - bc)}{\sqrt{([a+b+c+d]^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)([a+b+c+d]^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

$$\begin{aligned}
& A_{41}(j, k)^{(a,b=c=0,d)} \\
&= \frac{2(ad - 00)}{\sqrt{([a + 0 + 0 + d]^2 - a^2 - 0^2 - 0^2 - d^2 + 2a0 + 2d0)([a + 0 + 0 + d]^2 - a^2 - 0^2 - 0^2 - d^2 + 2a0 + 2d0)}} \\
&= \frac{2ad}{\sqrt{([a + d]^2 - a^2 - d^2)([a + d]^2 - a^2 - d^2)}} = \frac{2ad}{\sqrt{([a + d]^2 - a^2 - d^2)^2}} \\
&= \frac{2ad}{[a + d]^2 - a^2 - d^2} = \frac{2ad}{a^2 + 2ad + d^2 - a^2 - d^2} = \frac{2ad}{2ad} = 1 \text{ (máximo)}
\end{aligned}$$

**Kuder & Richardson**  $[-\infty, 1]$

$$\begin{aligned}
A_{42}(j, k)^{(a,b,c,d)} &= \frac{4(ad - bc)}{(a + b)(c + d) + (a + c)(b + d) + 2(ad - bc)} \\
A_{42}(j, k)^{(a,b=c=0,d)} &= \frac{4(ad - 0 * 0)}{(a + 0)(0 + d) + (a + 0)(0 + d) + 2(ad - 0 * 0)} = \frac{4ad}{ad + ad + 2(ad)} = \\
&= \frac{4ad}{4ad} = 1 \text{ (máximo)}
\end{aligned}$$

**Kulczynski\_1**  $[0, +\infty]$

$$\begin{aligned}
A_{43}(j, k)^{(a,b,c,d)} &= \frac{a}{b + c} \\
A_{43}(j, k)^{(a,b=c=0,d)} &= \frac{a}{0 + 0} = \frac{a}{0} \rightarrow +\infty \text{ (máximo)}
\end{aligned}$$

**Kulczynski\_2**  $[0, 1]$

$$\begin{aligned}
A_{44}(j, k)^{(a,b,c,d)} &= \frac{1}{2} \left( \frac{a}{a + b} + \frac{a}{a + c} \right) \\
A_{44}(j, k)^{(a,b=c=0,d)} &= \frac{1}{2} \left( \frac{a}{a + 0} + \frac{a}{a + 0} \right) = \frac{1}{2} \left( \frac{a}{a} + \frac{a}{a} \right) = \frac{1}{2} (1 + 1) = \frac{1}{2} * 2 = 1 \text{ (máximo)}
\end{aligned}$$

**Lamont & Grant**  $[0, 0.5]$

$$\begin{aligned}
A_{45}(j, k)^{(a,b,c,d)} &= \frac{a}{2a + b + c} \\
A_{45}(j, k)^{(a,b=c=0,d)} &= \frac{a}{2a + 0 + 0} = \frac{a}{2a} = \frac{1}{2} = 0.5 \text{ (máximo)}
\end{aligned}$$

**Legendre & Legendre**  $[0, 1]$

$$\begin{aligned}
A_{46}(j, k)^{(a,b,c,d)} &= \frac{3a}{3a + b + c} \\
A_{46}(j, k)^{(a,b=c=0,d)} &= \frac{3a}{3a + 0 + 0} = \frac{3a}{3a} = 1 \text{ (máximo)}
\end{aligned}$$

**Loevinger**  $[-\infty, 1]$

$$A_{47}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\min.\{(a+b)(b+d), (a+c)(c+d)\}}$$

$$A_{47}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{\min.\{(a+0)(0+d), (a+0)(0+d)\}} = \frac{ad}{\min.\{(ad), (ad)\}} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

**Maxwell & Pilliner**  $[-1, 1]$

$$A_{48}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(c+d) + (a+c)(b+d)}$$

$$A_{48}(j, k)^{(a,b=c=0,d)} = \frac{2(ad - 0 * 0)}{(a+0)(0+d) + (a+0)(0+d)} = \frac{2ad}{ad + ad} = \frac{2ad}{2ad} = 1 \text{ (máximo)}$$

**McConnaughey**  $[-1, 1]$

$$A_{49}(j, k)^{(a,b,c,d)} = \frac{a^2 - bc}{(a+b)(a+c)}$$

$$A_{49}(j, k)^{(a,b=c=0,d)} = \frac{a^2 - 0 * 0}{(a+0)(a+0)} = \frac{a^2}{aa} = \frac{a^2}{a^2} = 1 \text{ (máximo)}$$

**Michael**  $[-1, 1]$

$$A_{50}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a+d)^2 + (b+c)^2}$$

$$A_{50}(j, k)^{(a,b=c=0,d)} = \frac{4(ad - 0 * 0)}{(a+d)^2 + (0+0)^2} = \frac{4ad}{(a+d)^2} \text{ NO ES EL MÁXIMO}$$

$$\text{Si } a = d \rightarrow A_{50}(j, k)^{(a,b=c=0,d)} = \frac{4d^2}{(2d)^2} = \frac{4d^2}{4d^2} = 1 \text{ (máximo)}$$

**Mountford**  $[0, 2]$

$$A_{51}(j, k)^{(a,b,c,d)} = \frac{a}{\frac{1}{2}(ab + ac) + bc}$$

$$A_{51}(j, k)^{(a,b=c=0,d)} = \frac{a}{\frac{1}{2}(a0 + a0) + 0 * 0} = \frac{a}{\frac{1}{2}(0)} = \frac{a}{0} \rightarrow +\infty \text{ NO ES EL MÁXIMO}$$

**Ochiai**  $[0, 1]$

$$A_{52}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

$$A_{52}(j, k)^{(a,b=c=0,d)} = \frac{a}{\sqrt{(a+0)(a+0)}} = \frac{a}{\sqrt{aa}} = \frac{a}{\sqrt{a^2}} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Pearson & Heron** [-1, 1]

$$A_{53}(j, k)^{(a,b,c,d)} = \cos \frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{53}(j, k)^{(a,b=c=0,d)} = \cos \frac{180\sqrt{0*0}}{\sqrt{ad} + \sqrt{0*0}} = \cos \frac{180*0}{\sqrt{ad}} = \cos(0) = 1 \text{ (máximo)}$$

**Pearson\_1** [0, n]

$$A_{54}(j, k)^{(a,b,c,d)} = \chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$A_{54}(j, k)^{(a,b=c=0,d)} = \frac{(a+0+0+d)(ad - 0*0)^2}{(a+0)(0+d)(a+0)(0+d)} = \frac{(a+d)(ad)^2}{adad} = \frac{(a+d)(ad)^2}{(ad)^2} = a+d$$

$$n = a + b + c + d = a + 0 + 0 + d = a + d \rightarrow A_{54}(j, k)^{(a,b=c=0,d)} = a + d = n \text{ (máximo)}$$

**Pearson\_2** [0, 1]

$$A_{55}(j, k)^{(a,b,c,d)} = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{\frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{(a+b+c+d) + \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

$$\begin{aligned} A_{55}(j, k)^{(a,b=c=0,d)} &= \sqrt{\frac{\frac{(a+0+0+d)(ad - 0*0)^2}{(a+0)(0+d)(a+0)(0+d)}}{(a+0+0+d) + \frac{(a+0+0+d)(ad - 0*0)^2}{(a+0)(0+d)(a+0)(0+d)}}} \\ &= \sqrt{\frac{\frac{(a+d)(ad)^2}{adad}}{(a+d) + \frac{(a+d)(ad)^2}{adad}}} = \sqrt{\frac{\frac{(a+d)(ad)^2}{(ad)^2}}{(a+d) + \frac{(a+d)(ad)^2}{(ad)^2}}} \\ &= \sqrt{\frac{(a+d)}{(a+d) + (a+d)}} = \sqrt{\frac{(a+d)}{2(a+d)}} = \sqrt{\frac{1}{2}} \text{ NO ES EL MÁXIMO} \end{aligned}$$

**Peirce\_1** [0, 1]

$$A_{56}(j, k)^{(a,b,c,d)} = \frac{ab + bc}{ab + 2bc + cd}$$

$$A_{56}(j, k)^{(a,b=c=0,d)} = \frac{a0 + 0 * 0}{a0 + 2 * 0 * 0 + 0d} = \frac{0}{0} = \text{INDETERMINADO, NO ES EL MÁXIMO}$$

### Peirce\_2 [-1, 1]

$$A_{57}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + c)(b + d)}$$

$$A_{57}(j, k)^{(a,b=c=0,d)} = \frac{ad - 00}{(a + 0)(0 + d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

### Peirce\_3 [-1, 1]

$$A_{58}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + b)(c + d)}$$

$$A_{58}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{(a + 0)(0 + d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

### Roger & Tanimoto [0, 1]

$$A_{59}(j, k)^{(a,b,c,d)} = \frac{a + d}{a + 2(b + c) + d}$$

$$A_{59}(j, k)^{(a,b=c=0,d)} = \frac{a + d}{a + 2(0 + 0) + d} = \frac{a + d}{a + d} = 1 \text{ (máximo)}$$

### Rogot & Goldberg [0, 1]

$$A_{60}(j, k)^{(a,b,c,d)} = \frac{a}{2a + b + c} + \frac{d}{2d + b + c}$$

$$A_{60}(j, k)^{(a,b=c=0,d)} = \frac{a}{2a + 0 + 0} + \frac{d}{2d + 0 + 0} = \frac{a}{2a} + \frac{d}{2d} = \frac{1}{2} + \frac{1}{2} = 1 \text{ (máximo)}$$

### Russell & Rao [0, 1]

$$A_{61}(j, k)^{(a,b,c,d)} = \frac{a}{n} = \frac{a}{a + b + c + d}$$

$$A_{61}(j, k)^{(a,b=c=0,d)} = \frac{a}{a + 0 + 0 + d} = \frac{a}{a + d} < 1 \text{ NO ES EL MÁXIMO}$$

$$\text{Si } a = d \rightarrow A_{61}(j, k)^{(a,b=c=0,d)} = \frac{d}{d + d} = \frac{d}{2d} = \frac{1}{2} < 1 \text{ NO ES EL MÁXIMO}$$

**Scott** [-1, 1]

$$A_{62}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(2d + b + c)}$$

$$A_{62}(j, k)^{(a,b=c=0,d)} = \frac{4(ad - 00) - (0 - 0)^2}{(2a + 0 + 0)(2d + 0 + 0)} = \frac{4ad - 0}{(2a)(2d)} = \frac{4ad}{4ad} = 1 \text{ (máximo)}$$

**Simpson** [0, 1]

$$A_{63}(j, k)^{(a,b,c,d)} = \frac{a}{\text{mín.}\{(a + b), (a + c)\}}$$

$$A_{63}(j, k)^{(a,b=c=0,d)} = \frac{a}{\text{mín.}\{(a + 0), (a + 0)\}} = \frac{a}{\text{mín.}\{(a), (a)\}} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Sokal & Michener** [0, 1]

$$A_{64}(j, k)^{(a,b,c,d)} = \frac{a + d}{n} = \frac{a + d}{a + b + c + d}$$

$$A_{64}(j, k)^{(a,b=c=0,d)} = \frac{a + d}{a + 0 + 0 + d} = \frac{a + d}{a + d} = 1 \text{ (máximo)}$$

**Sokal & Sneath\_1** [0, 1]

$$A_{65}(j, k)^{(a,b,c,d)} = \frac{a}{a + 2(b + c)}$$

$$A_{65}(j, k)^{(a,b=c=0,d)} = \frac{a}{a + 2(0 + 0)} = \frac{a}{a + 0} = \frac{a}{a} = 1 \text{ (máximo)}$$

**Sokal & Sneath\_2** [0, 1]

$$A_{66}(j, k)^{(a,b,c,d)} = \frac{1}{4} \left( \frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{c + d} + \frac{d}{b + d} \right)$$

$$A_{66}(j, k)^{(a,b=c=0,d)} = \frac{1}{4} \left( \frac{a}{a + 0} + \frac{a}{a + 0} + \frac{d}{0 + d} + \frac{d}{0 + d} \right) = \frac{1}{4} \left( \frac{a}{a} + \frac{a}{a} + \frac{d}{d} + \frac{d}{d} \right) \\ = \frac{1}{4} (1 + 1 + 1 + 1) = \frac{4}{4} = 1 \text{ (máximo)}$$

**Sokal & Sneath\_3** [0, +∞]

$$A_{67}(j, k)^{(a,b,c,d)} = \frac{a + d}{b + c}$$

$$A_{67}(j, k)^{(a,b=c=0,d)} = \frac{a + d}{0 + 0} = \frac{a + d}{0} \rightarrow +\infty \text{ (máximo)}$$

**Sokal & Sneath\_4** [0, 1]

$$A_{68}(j, k)^{(a,b,c,d)} = \frac{a + d}{a + \frac{1}{2}(b + c) + d}$$

$$A_{68}(j, k)^{(a,b=c=0,d)} = \frac{a + d}{a + \frac{1}{2}(0 + 0) + d} = \frac{a + d}{a + 0 + d} = \frac{a + d}{a + d} = 1 \text{ (máximo)}$$

**Sokal & Sneath\_5** [0, 1]

$$A_{69}(j, k)^{(a,b,c,d)} = \frac{ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

$$A_{69}(j, k)^{(a,b=c=0,d)} = \frac{ad}{\sqrt{(a + 0)(0 + d)(a + 0)(0 + d)}} = \frac{ad}{\sqrt{adad}} = \frac{ad}{\sqrt{(ad)^2}} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

**Sorensen** [0, 1]

$$A_{70}(j, k)^{(a,b,c,d)} = \frac{2a}{2a + b + c}$$

$$A_{70}(j, k)^{(a,b=c=0,d)} = \frac{2a}{2a + 0 + 0} = \frac{2a}{2a} = 1 \text{ (máximo)}$$

**Sorgenfrei** [0, 1]

$$A_{71}(j, k)^{(a,b,c,d)} = \frac{a^2}{(a + b)(a + c)}$$

$$A_{71}(j, k)^{(a,b=c=0,d)} = \frac{a^2}{(a + 0)(a + 0)} = \frac{a^2}{aa} = \frac{a^2}{a^2} = 1 \text{ (máximo)}$$

**Stiles**  $[-\infty, +\infty]$ 

$$A_{72}(j, k)^{(a,b,c,d)} = \log \frac{n(|ad - bc| - \frac{n}{2})^2}{(a + b)(a + c)(b + d)(c + d)}$$

$$= \log \frac{(a + b + c + d)(|ad - bc| - \frac{a + b + c + d}{2})^2}{(a + b)(a + c)(b + d)(c + d)}$$

$$\begin{aligned}
A_{72}(j, k)^{(a,b=c=0,d)} &= \log \frac{(a+0+0+d) \left( |ad-0*0| - \frac{a+0+0+d}{2} \right)^2}{(a+0)(a+0)(0+d)(0+d)} \\
&= \log \frac{(a+d) \left( ad - \frac{a+d}{2} \right)^2}{aadd} = \log \frac{(a+d) \left( \frac{2ad - (a+d)}{2} \right)^2}{(ad)^2} \\
&= \log \frac{(a+d)(2ad - (a+d))^2}{(2ad)^2} \quad \text{NO ES EL MÁXIMO}
\end{aligned}$$

**Tarwid** [-1, 1]

$$A_{73}(j, k)^{(a,b,c,d)} = \frac{na - (a+b)(a+c)}{na + (a+b)(a+c)} = \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d)a + (a+b)(a+c)}$$

$$\begin{aligned}
A_{73}(j, k)^{(a,b=c=0,d)} &= \frac{(a+0+0+d)a - (a+0)(a+0)}{(a+0+0+d)a + (a+0)(a+0)} = \frac{(a+d)a - aa}{(a+d)a + aa} = \frac{aa + ad - aa}{aa + ad + aa} \\
&= \frac{ad}{2aa + ad} = \frac{ad}{2a^2 + ad} = \frac{a(d)}{a(2a+d)} = \frac{d}{2a+d} = \frac{d}{d+(2a)} \\
&< 1 \quad \text{NO ES EL MÁXIMO}
\end{aligned}$$

**Tschuproff** [0, 1]

$$A_{74}(j, k)^{(a,b,c,d)} = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}} = \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \cdot (a+b+c+d)}}$$

$$\begin{aligned}
A_{74}(j, k)^{(a,b=c=0,d)} &= \sqrt{\frac{(a+0+0+d)(ad-0*0)^2}{(a+0)(0+d)(a+0)(0+d)}} = \sqrt{\frac{(a+d)(ad)^2}{adad}} = \sqrt{\frac{(a+d)(ad)^2}{(ad)^2}} \\
&= \sqrt{\frac{a+d}{a+d}} = \sqrt{1} = 1 \quad (\text{máximo})
\end{aligned}$$

**Yule \_1** [-1, 1]

$$A_{75}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \phi$$

$$\begin{aligned}
A_{75}(j, k)^{(a,b=c=0,d)} &= \frac{ad - 0*0}{\sqrt{(a+0)(0+d)(a+0)(0+d)}} = \frac{ad}{\sqrt{adad}} = \frac{ad}{\sqrt{a^2d^2}} = \frac{ad}{ad} \\
&= 1 \quad (\text{máximo})
\end{aligned}$$

**Yule \_2** [-1, 1]

$$A_{76}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{ad + bc}$$

$$A_{76}(j, k)^{(a,b=c=0,d)} = \frac{ad - 0 * 0}{ad + 0 * 0} = \frac{ad}{ad} = 1 \text{ (m\u00e1ximo)}$$

**Yule \_3** [-1, 1]

$$A_{77}(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{77}(j, k)^{(a,b=c=0,d)} = \frac{\sqrt{ad} - \sqrt{0 * 0}}{\sqrt{ad} + \sqrt{0 * 0}} = \frac{\sqrt{ad}}{\sqrt{ad}} = 1 \text{ (m\u00e1ximo)}$$



### ANEXO 3

Propiedad 4: Asociación completa o absoluta

Si  $\max. A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} > A_i(j, k)^{(a,b=0,c,d)}$  y  $A_i(j, k)^{(a,b=0,c,d)} > A_i(j, k)^{(a,b,c=0,d)} \rightarrow A_i(j, k)$  es un coeficiente de Asociación absoluta

Si  $\max. A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a,b=c=0,d)} = A_i(j, k)^{(a,b=0,c,d)} = A_i(j, k)^{(a,b,c=0,d)} \rightarrow A_i(j, k)$  es un coeficiente de Asociación completa

**Anónimo\_1** [-1, 1]

$$A_1(j, k)^{(a,b,c,d)} = \frac{2a - (b + c)}{2a + (b + c)}$$

$$A_1(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_1(j, k)^{(a,b=0,c,d)} = \frac{2a - (0 + c)}{2a + (0 + c)} = \frac{2a - c}{2a + c} < 1$$

$$A_1(j, k)^{(a,b,c=0,d)} = \frac{2a - (b + 0)}{2a + (b + 0)} = \frac{2a - b}{2a + b} < 1$$

$\max. A_1(j, k)^{(a,b,c,d)} = A_1(j, k)^{(a,b=c=0,d)} > A_1(j, k)^{(a,b=0,c,d)}$  y  $A_1(j, k)^{(a,b=0,c,d)} > A_1(j, k)^{(a,b,c=0,d)} \rightarrow A_1(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Anónimo\_2** [0, 1]

$$A_2(j, k)^{(a,b,c,d)} = \frac{2d}{2d + b + c}$$

$$A_2(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_2(j, k)^{(a,b=0,c,d)} = \frac{2d}{2d + 0 + c} = \frac{2d}{2d + c} < 1$$

$$A_2(j, k)^{(a,b,c=0,d)} = \frac{2d}{2d + b + 0} = \frac{2d}{2d + b} < 1$$

$\max. A_2(j, k)^{(a,b,c,d)} = A_2(j, k)^{(a,b=c=0,d)} > A_2(j, k)^{(a,b=0,c,d)}$  y  $A_2(j, k)^{(a,b=0,c,d)} > A_2(j, k)^{(a,b,c=0,d)} \rightarrow A_2(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Anónimo\_3** [-1, 1]

$$A_3(j, k)^{(a,b,c,d)} = \frac{2d - (b + c)}{2d + (b + c)}$$

$$A_3(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_3(j, k)^{(a,b=0,c,d)} = \frac{2d - (0 + c)}{2d + (0 + c)} = \frac{2d - c}{2d + c} < 1$$

$$A_3(j, k)^{(a,b,c=0,d)} = \frac{2d - (b + 0)}{2d + (b + 0)} = \frac{2d - b}{2d + b} < 1$$

máx.  $A_3(j, k)^{(a,b,c,d)} = A_3(j, k)^{(a,b=c=0,d)} > A_3(j, k)^{(a,b=0,c,d)}$  y  $> A_3(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_3(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Anónimo\_4** [0, 1]

$$A_4(j, k)^{(a,b,c,d)} = \frac{4ad}{4ad + (a + d)(b + c)}$$

$$A_4(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_4(j, k)^{(a,b=0,c,d)} = \frac{4ad}{4ad + (a + d)(0 + c)} = \frac{4ad}{4ad + (a + d)c} < 1$$

$$A_4(j, k)^{(a,b,c=0,d)} = \frac{4ad}{4ad + (a + d)(b + 0)} = \frac{4ad}{4ad + (a + d)b} < 1$$

máx.  $A_4(j, k)^{(a,b,c,d)} = A_4(j, k)^{(a,b=c=0,d)} > A_4(j, k)^{(a,b=0,c,d)}$  y  $> A_4(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_4(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Anónimo\_5** [-1, +∞]

$$A_5(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a + b)(a + c), (c + d)(b + d)\}}$$

$$A_5(j, k)^{(a,b=c=0,d)} \rightarrow \text{NO ES EL MÁXIMO}$$

**Baroni-Urbani & Buser\_1** [0, 1]

$$A_6(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

$$A_6(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_6(j, k)^{(a,b=0,c,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + 0 + c} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + c} < 1$$

$$A_6(j, k)^{(a,b,c=0,d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + 0} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b} < 1$$

máx.  $A_6(j, k)^{(a,b,c,d)} = A_6(j, k)^{(a,b=c=0,d)} > A_6(j, k)^{(a,b=0,c,d)}$  y  $> A_6(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_6(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Baroni-Urbani & Buser\_2** [-1, 1]

$$A_7(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} + a - b - c}{\sqrt{ad} + a + b + c}$$

$$A_7(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_7(j, k)^{(a,b=0,c,d)} = \frac{\sqrt{ad} + a - 0 - c}{\sqrt{ad} + a + 0 + c} = \frac{\sqrt{ad} + a - c}{\sqrt{ad} + a + c} < 1$$

$$A_7(j, k)^{(a,b,c=0,d)} = \frac{\sqrt{ad} + a - b - 0}{\sqrt{ad} + a + b + 0} = \frac{\sqrt{ad} + a - b}{\sqrt{ad} + a + b} < 1$$

máx.  $A_7(j, k)^{(a,b,c,d)} = A_7(j, k)^{(a,b=c=0,d)} > A_7(j, k)^{(a,b=0,c,d)}$  y  $> A_7(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_7(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

### Braun-Blanquet [0, 1]

$$A_8(j, k)^{(a,b,c,d)} = \frac{a}{\text{máx.}\{(a+b), (a+c)\}}$$

$$A_8(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_8(j, k)^{(a,b=0,c,d)} = \frac{a}{\text{máx.}\{(a+0), (a+c)\}} = \frac{a}{\text{máx.}\{(a), (a+c)\}} = \frac{a}{a+c} < 1$$

$$A_8(j, k)^{(a,b,c=0,d)} = \frac{a}{\text{máx.}\{(a+b), (a+0)\}} = \frac{a}{\text{máx.}\{(a+b), (a)\}} = \frac{a}{a+b} < 1$$

máx.  $A_8(j, k)^{(a,b,c,d)} = A_8(j, k)^{(a,b=c=0,d)} > A_8(j, k)^{(a,b=0,c,d)}$  y  $> A_8(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_8(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

### Clement [0, n]

$$A_9(j, k)^{(a,b,c,d)} = \frac{a(c+d)}{(a+b)} + \frac{d(a+b)}{(c+d)}$$

$$A_9(j, k)^{(a,b=c=0,d)} = n \text{ (máximo)}$$

$$A_9(j, k)^{(a,b=0,c,d)} = \frac{a(c+d)}{(a+0)} + \frac{d(a+0)}{(c+d)} = \frac{a(c+d)}{a} + \frac{da}{(c+d)} = (c+d) + \frac{ad}{c+d}$$

$$n = a + b + c + d = a + 0 + c + d = a + c + d \rightarrow c + d = n - a$$

$$A_9(j, k)^{(a,b=0,c,d)} = (n - a) + \frac{ad}{(n - a)} < n$$

$$A_9(j, k)^{(a,b,c=0,d)} = \frac{a(0+d)}{(a+b)} + \frac{d(a+b)}{(0+d)} = \frac{ad}{(a+b)} + \frac{d(a+b)}{d} = \frac{ad}{a+b} + (a+b)$$

$$n = a + b + c + d = a + b + 0 + d = a + b + d \rightarrow a + b = n - d$$

$$A_9(j, k)^{(a,b,c=0,d)} = (n - d) + \frac{ad}{(n - d)} < n$$

máx.  $A_9(j, k)^{(a,b,c,d)} = A_9(j, k)^{(a,b=c=0,d)} > A_9(j, k)^{(a,b=0,c,d)}$  y  $> A_9(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_9(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Clifford & Stephenson** [0, 1]

$$A_{10}(j, k)^{(a,b,c,d)} = \frac{\chi^2}{n} = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$A_{10}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{10}(j, k)^{(a,b=0,c,d)} &= \frac{(ad - 0c)^2}{(a+0)(c+d)(a+c)(0+d)} = \frac{(ad)^2}{a(c+d)(a+c)d} = \frac{(ad)^2}{(ad)(c+d)(a+c)} \\ &= \frac{ad}{(c+d)(a+c)} = \frac{ad}{ac + c^2 + ad + cd} = \frac{ad}{ad + (ac + cd + c^2)} < 1 \end{aligned}$$

$$\begin{aligned} A_{10}(j, k)^{(a,b,c=0,d)} &= \frac{(ad - b0)^2}{(a+b)(0+d)(a+0)(b+d)} = \frac{(ad)^2}{(a+b)da(b+d)} \\ &= \frac{(ad)^2}{(ad)(a+b)(b+d)} = \frac{ad}{(a+b)(b+d)} = \frac{ad}{ab + ad + b^2 + bd} \\ &= \frac{ad}{ad + (ab + bd + b^2)} < 1 \end{aligned}$$

máx.  $A_{10}(j, k)^{(a,b,c,d)} = A_{10}(j, k)^{(a,b=c=0,d)} > A_{10}(j, k)^{(a,b=0,c,d)}$  y  $> A_{10}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{10}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Cohen** [-1, 1]

$$A_{11}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(b+d) + (a+c)(c+d)}$$

$$A_{11}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{11}(j, k)^{(a,b=0,c,d)} &= \frac{2(ad - 0c)}{(a+0)(0+d) + (a+c)(c+d)} = \frac{2ad}{ad + (a+c)(c+d)} \\ &= \frac{2ad}{2ad} = \frac{2ad}{ad + (ac + ad + c^2 + cd)} = \frac{2ad}{2ad + (ac + cd + c^2)} < 1 \end{aligned}$$

$$\begin{aligned} A_{11}(j, k)^{(a,b,c=0,d)} &= \frac{2(ad - b0)}{(a+b)(b+d) + (a+0)(0+d)} = \frac{2ad}{(a+b)(b+d) + ad} \\ &= \frac{2ad}{2ad} = \frac{2ad}{ad + (ab + ad + b^2 + bd)} = \frac{2ad}{2ad + (ab + bd + b^2)} < 1 \end{aligned}$$

máx.  $A_{11}(j, k)^{(a,b,c,d)} = A_{11}(j, k)^{(a,b=c=0,d)} > A_{11}(j, k)^{(a,b=0,c,d)}$  y  $> A_{11}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{11}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Cole\_1** [-1, 1]

$$A_{12}(j, k)^{(a,b,c,d)} = \frac{\sqrt{2}(ad - bc)}{\sqrt{(ad - bc)^2 + (a+b)(c+d)(a+c)(b+d)}}$$

$$A_{12}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{12}(j, k)^{(a,b=0,c,d)} &= \frac{\sqrt{2}(ad - 0c)}{\sqrt{(ad - 0c)^2 + (a + 0)(c + d)(a + c)(0 + d)}} \\ &= \frac{\sqrt{2}ad}{\sqrt{(ad)^2 + a(c + d)(a + c)d}} = \frac{\sqrt{2}ad}{\sqrt{(ad)^2 + ad(c + d)(a + c)}} \\ &= \frac{\sqrt{2}ad}{\sqrt{(ad)[ad + (c + d)(a + c)]}} = \frac{\sqrt{2}ad}{\sqrt{ad}\sqrt{ad + (ac + c^2 + ad + cd)}} \\ &= \frac{\sqrt{2}\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{2ad + ac + cd + c^2}} = \frac{\sqrt{2}\sqrt{ad}}{\sqrt{2ad + (ac + cd + c^2)}} \\ &= \frac{\sqrt{2ad}}{\sqrt{2ad + (ac + cd + c^2)}} < 1 \end{aligned}$$

$$\begin{aligned} A_{12}(j, k)^{(a,b,c=0,d)} &= \frac{\sqrt{2}(ad - b0)}{\sqrt{(ad - b0)^2 + (a + b)(0 + d)(a + 0)(b + d)}} \\ &= \frac{\sqrt{2}ad}{\sqrt{(ad)^2 + (a + b)da(b + d)}} = \frac{\sqrt{2}ad}{\sqrt{(ad)^2 + ad(a + b)(b + d)}} \\ &= \frac{\sqrt{2}ad}{\sqrt{(ad)[ad + (a + b)(b + d)]}} = \frac{\sqrt{2}ad}{\sqrt{ad}\sqrt{ad + (ab + ad + b^2 + bd)}} \\ &= \frac{\sqrt{2}\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{2ad + ab + bd + b^2}} = \frac{\sqrt{2}\sqrt{ad}}{\sqrt{2ad + (ab + bd + b^2)}} \\ &= \frac{\sqrt{2ad}}{\sqrt{2ad + (ab + bd + b^2)}} < 1 \end{aligned}$$

máx.  $A_{12}(j, k)^{(a,b,c,d)} = A_{12}(j, k)^{(a,b=c=0,d)} > A_{12}(j, k)^{(a,b=0,c,d)}$  y  $> A_{12}(j, k)^{(a,b,c=0,d)}$  —  
 $\rightarrow A_{12}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Cole\_2** [0, 1]

$$A_{13}(j, k)^{(a,b,c,d)} = \frac{\Phi}{\Phi_{\text{máx.}}} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} \cdot \frac{\Phi_{\text{máx.}}}{\Phi_{\text{máx.}}}$$

$$A_{13}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{13}(j, k)^{(a,b=0,c,d)} &= \frac{ad - 0c}{\sqrt{(a + 0)(c + d)(a + c)(0 + d)}} = \frac{ad}{\sqrt{a(c + d)(a + c)d}} \\ &= \frac{ad}{\sqrt{ad(c + d)(a + c)}} = \frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(c + d)(a + c)}} = \frac{\sqrt{ad}}{\sqrt{ac + c^2 + ad + cd}} \\ &= \frac{\sqrt{ad}}{\sqrt{ad + (ac + cd + c^2)}} \\ &= \frac{\sqrt{ad}}{\Phi_{\text{máx.}}} \end{aligned}$$

$$\text{Si } \Phi_{\text{máx.}} = 1 \rightarrow A_{13}(j, k)^{(a,b=0,c,d)} = \frac{\sqrt{ad}}{\sqrt{ad + (ac + cd + c^2)}} = \frac{\sqrt{ad}}{\sqrt{ad + (ac + cd + c^2)}} < 1$$

$$\begin{aligned} A_{13}(j, k)^{(a,b,c=0,d)} &= \frac{\frac{ad - b0}{\sqrt{(a+b)(0+d)(a+0)(b+d)}}}{\Phi_{\text{máx.}}} = \frac{\frac{ad}{\sqrt{(a+b)da(b+d)}}}{\Phi_{\text{máx.}}} \\ &= \frac{\frac{ad}{\sqrt{ad(a+b)(b+d)}}}{\Phi_{\text{máx.}}} = \frac{\frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(a+b)(b+d)}}}{\Phi_{\text{máx.}}} = \frac{\frac{\sqrt{ad}}{\sqrt{ab + ad + b^2 + bd}}}{\Phi_{\text{máx.}}} \\ &= \frac{\frac{\sqrt{ad}}{\sqrt{ad + (ab + bd + b^2)}}}{\Phi_{\text{máx.}}} \end{aligned}$$

$$\text{Si } \Phi_{\text{máx.}} = 1 \rightarrow A_{13}(j, k)^{(a,b,c=0,d)} = \frac{\sqrt{ad}}{\sqrt{ad + (ab + bd + b^2)}} = \frac{\sqrt{ad}}{\sqrt{ad + (ab + bd + b^2)}} < 1$$

máx.  $A_{13}(j, k)^{(a,b,c,d)} = A_{13}(j, k)^{(a,b=c=0,d)} > A_{13}(j, k)^{(a,b=0,c,d)}$  y  $> A_{13}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{13}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Cole\_3**  $[-\infty, 1]$

$$A_{14}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+b)(b+d)}$$

$$A_{14}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{14}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{(a+0)(0+d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{14}(j, k)^{(a,b,c=0,d)} &= \frac{ad - b0}{(a+b)(b+d)} = \frac{ad}{(a+b)(b+d)} = \frac{ad}{ab + ad + b^2 + bd} \\ &= \frac{ad}{ad + (ab + bd + b^2)} < 1 \end{aligned}$$

máx.  $A_{14}(j, k)^{(a,b,c,d)} = A_{14}(j, k)^{(a,b=c=0,d)} = A_{14}(j, k)^{(a,b=0,c,d)}$  y  $> A_{14}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{14}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA (b = 0) Y ABSOLUTA (c = 0)

**Cole\_4**  $[-\infty, 1]$

$$A_{15}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+c)(c+d)}$$

$$A_{15}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{15}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{(a+c)(c+d)} = \frac{ad}{(a+c)(c+d)} = \frac{ad}{ac + ad + c^2 + cd}$$

$$= \frac{ad}{ad + (ac + cd + c^2)} < 1$$

$$A_{15}(j, k)^{(a,b,c=0,d)} = \frac{ad - b0}{(a+0)(0+d)} = \frac{ad}{ad} = 1 \text{ (máximo)}$$

máx.  $A_{15}(j, k)^{(a,b,c,d)} = A_{15}(j, k)^{(a,b=c=0,d)} > A_{15}(j, k)^{(a,b=0,c,d)}$  y  $= A_{15}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{15}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA ( $b = 0$ ) Y COMPLETA ( $c = 0$ )

**Dennis**  $[-\infty, +\infty]$

$$A_{16}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a+b+c+d)(a+b)(a+c)}}$$

$A_{16}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Dice\_1**  $[0, 1]$

$$A_{17}(j, k)^{(a,b,c,d)} = \frac{a}{a+b}$$

$A_{17}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{17}(j, k)^{(a,b=0,c,d)} = \frac{a}{a+0} = \frac{a}{a} = 1 \text{ (máximo)}$$

$$A_{17}(j, k)^{(a,b,c=0,d)} = \frac{a}{a+b} < 1$$

máx.  $A_{17}(j, k)^{(a,b,c,d)} = A_{17}(j, k)^{(a,b=c=0,d)} = A_{17}(j, k)^{(a,b=0,c,d)}$  y  $> A_{17}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{17}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA ( $b = 0$ ) Y ABSOLUTA ( $c = 0$ )

**Dice\_2**  $[0, 1]$

$$A_{18}(j, k)^{(a,b,c,d)} = \frac{a}{a+c}$$

$A_{18}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{18}(j, k)^{(a,b=0,c,d)} = \frac{a}{a+c} < 1$$

$$A_{18}(j, k)^{(a,b,c=0,d)} = \frac{a}{a+0} = \frac{a}{a} = 1 \text{ (máximo)}$$

máx.  $A_{18}(j, k)^{(a,b,c,d)} = A_{18}(j, k)^{(a,b=c=0,d)} > A_{18}(j, k)^{(a,b=0,c,d)}$  y  $= A_{18}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{18}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA ( $b = 0$ ) Y COMPLETA ( $c = 0$ )

**Dice\_3** [0, 2]

$$A_{19}(j, k)^{(a,b,c,d)} = \frac{2a}{(a+b)(a+c)}$$

$$A_{19}(j, k)^{(a,b=c=0,d)} = 2 \text{ (máximo) Si } a = 1$$

$$A_{19}(j, k)^{(a,b=0,c,d)} = \frac{2 * 1}{(1+0)(1+c)} = \frac{2}{1+c} < 2$$

$$A_{19}(j, k)^{(a,b,c=0,d)} = \frac{2 * 1}{(1+b)(1+0)} = \frac{2}{1+b} < 2$$

máx.  $A_{19}(j, k)^{(a,b,c,d)} = A_{19}(j, k)^{(a=1,b=c=0,d)} > A_{19}(j, k)^{(a,b=0,c,d)}$  y  $> A_{19}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{19}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA Si  $a = 1$

**Digby** [-1, 1]

$$A_{20}(j, k)^{(a,b,c,d)} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}}$$

$$A_{20}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{20}(j, k)^{(a,b=0,c,d)} = \frac{(ad)^{3/4} - (0c)^{3/4}}{(ad)^{3/4} + (0c)^{3/4}} = \frac{(ad)^{3/4}}{(ad)^{3/4}} = 1 \text{ (máximo)}$$

$$A_{20}(j, k)^{(a,b,c=0,d)} = \frac{(ad)^{3/4} - (b0)^{3/4}}{(ad)^{3/4} + (b0)^{3/4}} = \frac{(ad)^{3/4}}{(ad)^{3/4}} = 1 \text{ (máximo)}$$

máx.  $A_{20}(j, k)^{(a,b,c,d)} = A_{20}(j, k)^{(a,b=c=0,d)} = A_{20}(j, k)^{(a,b=0,c,d)}$  y  $= A_{20}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{20}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA

**Doolittle** [0, 1]

$$A_{21}(j, k)^{(a,b,c,d)} = \Phi^2 = \frac{(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$A_{21}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{21}(j, k)^{(a,b=0,c,d)} &= \frac{(ad - 0c)^2}{(a+0)((c+d)(a+c)(0+d))} = \frac{(ad)^2}{a(c+d)(a+c)d} = \frac{(ad)^2}{ad(c+d)(a+c)} \\ &= \frac{ad}{(c+d)(a+c)} = \frac{ad}{ac + c^2 + ad + cd} = \frac{ad}{ad + (ac + cd + c^2)} < 1 \end{aligned}$$

$$\begin{aligned} A_{21}(j, k)^{(a,b,c=0,d)} &= \frac{(ad - b0)^2}{(a+b)((0+d)(a+0)(b+d))} = \frac{(ad)^2}{(a+b)da(b+d)} = \frac{(ad)^2}{ad(a+b)(b+d)} \\ &= \frac{ad}{(a+b)(b+d)} = \frac{ad}{ab + ad + b^2 + bd} = \frac{ad}{ad + (ab + bd + b^2)} < 1 \end{aligned}$$

máx.  $A_{21}(j, k)^{(a,b,c,d)} = A_{21}(j, k)^{(a,b=c=0,d)} > A_{21}(j, k)^{(a,b=0,c,d)}$  y  $> A_{21}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{21}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Eyraud** [-1, 0]

$$A_{22}(j, k)^{(a,b,c,d)} = \frac{a - (a+b)(a+c)}{(a+b)(c+d)(a+c)(b+d)}$$

$A_{22}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Fager & McGowan** [-1, 1]

$$A_{23}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a+b)(a+c)}} - \frac{1}{2\sqrt{\text{máx.}\{(a+b), (a+c)\}}}$$

$A_{23}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Faith** [0, 1]

$$A_{24}(j, k)^{(a,b,c,d)} = \frac{a + \frac{d}{2}}{n} = \frac{a + \frac{d}{2}}{a + b + c + d}$$

$A_{24}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Fleiss**  $[-\infty, 1]$

$$A_{25}(j, k)^{(a,b,c,d)} = \frac{(ad - bc)[(a+b)(b+d) + (a+c)(c+d)]}{2(a+b)(a+c)(b+d)(c+d)}$$

$A_{25}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$\begin{aligned} A_{25}(j, k)^{(a,b=0,c,d)} &= \frac{(ad - 0c)[(a+0)(0+d) + (a+c)(c+d)]}{2(a+0)(a+c)(0+d)(c+d)} \\ &= \frac{(ad)[ad + (a+c)(c+d)]}{2a(a+c)d(c+d)} = \frac{(ad)[ad + (a+c)(c+d)]}{2ad(a+c)(c+d)} \\ &= \frac{ad + (a+c)(c+d)}{2(a+c)(c+d)} < 1 \end{aligned}$$

$$\begin{aligned} A_{25}(j, k)^{(a,b,c=0,d)} &= \frac{(ad - b0)[(a+b)(b+d) + (a+0)(0+d)]}{2(a+b)(a+0)(b+d)(0+d)} \\ &= \frac{(ad)[(a+b)(b+d) + ad]}{2(a+b)a(b+d)d} = \frac{(ad)[(a+b)(d+d) + ad]}{2ad(a+b)(b+d)} \\ &= \frac{(a+b)(b+d) + ad}{2(a+b)(b+d)} < 1 \end{aligned}$$

$\text{máx. } A_{25}(j, k)^{(a,b,c,d)} = A_{25}(j, k)^{(a,b=c=0,d)} > A_{25}(j, k)^{(a,b=0,c,d)}$  y  $> A_{25}(j, k)^{(a,b,c=0,d)}$  —  
 $\rightarrow A_{25}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Forbes\_1** [0,  $+\infty$ ]

$$A_{26}(j, k)^{(a,b,c,d)} = \frac{n * a}{(a+b)(a+c)} = \frac{(a+b+c+d)a}{(a+b)(a+c)}$$

$A_{26}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Forbes\_2** [-1, 1]

$$\begin{aligned} A_{27}(j, k)^{(a,b,c,d)} &= \frac{n * a - (a+b)(a+c)}{n * \text{máx.}\{(a+b), (a+c)\} - (a+b)(a+c)} \\ &= \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d) * \text{máx.}\{(a+b), (a+c)\} - (a+b)(a+c)} \end{aligned}$$

$A_{27}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$\begin{aligned} A_{27}(j, k)^{(a,b=0,c,d)} &= \frac{(a+0+c+d)a - (a+0)(a+c)}{(a+0+c+d) * \text{máx.}\{(a+0), (a+c)\} - (a+0)(a+c)} \\ &= \frac{(a+c+d)a - (a+c)a}{(a+c+d) * \text{máx.}\{(a), (a+c)\} - a(a+c)} \\ &= \frac{(a+c+d)a - a(a+c)}{(a+c+d)(a+c) - a(a+c)} = \frac{a(a+c+d) - a(a+c)}{(a+c+d)a + (a+c+d)c - a(a+c)} \\ &= \frac{a(a+c+d) - a(a+c)}{a(a+c+d) - a(a+c) + [c(a+c+d)]} < 1 \end{aligned}$$

$$\begin{aligned} A_{27}(j, k)^{(a,b,c=0,d)} &= \frac{(a+b+0+d)a - (a+b)(a+0)}{(a+b+0+d) * \text{máx.}\{(a+b), (a+0)\} - (a+b)(a+0)} \\ &= \frac{(a+b+d)a - (a+b)a}{(a+b+d) * \text{máx.}\{(a+b), (a)\} - (a+b)a} \\ &= \frac{(a+b+d)a - a(a+b)}{(a+c+d)(a+b) - a(a+b)} = \frac{a(a+b+d) - a(a+b)}{(a+b+d)a + (a+b+d)b - a(a+b)} \\ &= \frac{a(a+b+d) - a(a+b)}{a(a+b+d) - a(a+b) + [b(a+b+d)]} < 1 \end{aligned}$$

$\text{máx. } A_{27}(j, k)^{(a,b,c,d)} = A_{27}(j, k)^{(a,b=c=0,d)} > A_{27}(j, k)^{(a,b=0,c,d)} > A_{27}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{27}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Forbes\_3** [-1, +∞]

$$A_{28}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+b)(a+c)}$$

$A_{28}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Fossum** [0, +∞]

$$A_{29}(j, k)^{(a,b,c,d)} = \frac{n \left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)} = \frac{(a+b+c+d) \left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)}$$

$A_{29}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Gilbert & Wells**  $[-\infty, +\infty]$

$$A_{30}(j, k)^{(a,b,c,d)} = \log(a) - \log(n) - \log\left(\frac{a+b}{n}\right) - \log\left(\frac{a+c}{n}\right) \\ = \log(a) - \log(a+b+c+d) - \log\left(\frac{a+b}{a+b+c+d}\right) - \log\left(\frac{a+c}{a+b+c+d}\right)$$

$A_{30}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Goodall**  $[0, 1]$

$$A_{31}(j, k)^{(a,b,c,d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{n}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+c+d}}$$

$A_{31}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{31}(j, k)^{(a,b=0,c,d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+0+c+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+c+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{(a+d)}{(a+d)+c}} \\ = \frac{2}{\pi} \arcsin \sqrt{(\text{expresión} < 1)} = \frac{2}{\pi} \left( \text{expresión} < \frac{\pi}{2} \right) < 1$$

$$A_{31}(j, k)^{(a,b,c=0,d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+0+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{(a+d)}{(a+d)+b}} \\ = \frac{2}{\pi} \arcsin \sqrt{(\text{expresión} < 1)} = \frac{2}{\pi} \left( \text{expresión} < \frac{\pi}{2} \right) < 1$$

máx.  $A_{31}(j, k)^{(a,b,c,d)} = A_{31}(j, k)^{(a,b=c=0,d)} > A_{31}(j, k)^{(a,b=0,c,d)}$  y  $> A_{31}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{31}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Goodman & Kruskal**  $[-1, 1]$

$$A_{32}(j, k)^{(a,b,c,d)} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

$A_{32}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{32}(j, k)^{(a,b=0,c,d)} = \frac{2\text{mín.}\{(a), (d)\} - 0 - c}{2\text{mín.}\{(a), (d)\} + 0 + c} = \frac{2\text{mín.}\{(a), (d)\} - c}{2\text{mín.}\{(a), (d)\} + c} < 1$$

$$A_{32}(j, k)^{(a,b,c=0,d)} = \frac{2\text{mín.}\{(a), (d)\} - b - 0}{2\text{mín.}\{(a), (d)\} + b + 0} = \frac{2\text{mín.}\{(a), (d)\} - b}{2\text{mín.}\{(a), (d)\} + b} < 1$$

máx.  $A_{32}(j, k)^{(a,b,c,d)} = A_{32}(j, k)^{(a,b=c=0,d)} > A_{32}(j, k)^{(a,b=0,c,d)}$  y  $> A_{32}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{32}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Hamann**  $[-1, 1]$

$$A_{33}(j, k)^{(a,b,c,d)} = \frac{a + d - b - c}{n} = \frac{a + d - b - c}{a + d + b + c}$$

$$A_{33}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{33}(j, k)^{(a,b=0,c,d)} = \frac{a + d - 0 - c}{a + d + 0 + c} = \frac{a + d - c}{a + d + c} < 1$$

$$A_{33}(j, k)^{(a,b,c=0,d)} = \frac{a + d - b - 0}{a + d + b + 0} = \frac{a + d - b}{a + d + b} < 1$$

$$\text{máx. } A_{33}(j, k)^{(a,b,c,d)} = A_{33}(j, k)^{(a,b=c=0,d)} > A_{33}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{33}(j, k)^{(a,b,c=0,d)} - \\ \rightarrow A_{33}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA}$$

**Harris & Lahey** [0, n]

$$A_{34}(j, k)^{(a,b,c,d)} = \frac{a(2d + b + c)}{2(a + b + c)} + \frac{d(2a + b + c)}{2(d + b + c)}$$

$$A_{34}(j, k)^{(a,b=c=0,d)} = n \text{ (máximo)}$$

$$A_{34}(j, k)^{(a,b=0,c,d)} = \frac{a(2d + 0 + c)}{2(a + 0 + c)} + \frac{d(2a + 0 + c)}{2(d + 0 + c)} = \frac{a(2d + c)}{2(a + c)} + \frac{d(2a + c)}{2(d + c)} \\ = \frac{a[d + (c + d)]}{2(a + c)} + \frac{d[a + (a + c)]}{2(c + d)}$$

$$n = a + b + c + d = a + 0 + c + d = a + c + d \rightarrow a + c = n - d \text{ \& } c + d = n - a$$

$$A_{34}(j, k)^{(a,b=0,c,d)} = \frac{a[d + (n - a)]}{2(n - d)} + \frac{d[a + (n - d)]}{2(n - a)} < n$$

$$A_{34}(j, k)^{(a,b,c=0,d)} = \frac{a(2d + b + 0)}{2(a + b + 0)} + \frac{d(2a + b + 0)}{2(d + b + 0)} = \frac{a(2d + b)}{2(a + b)} + \frac{d(2a + b)}{2(d + b)} \\ = \frac{a[d + (b + d)]}{2(a + b)} + \frac{d[a + (a + b)]}{2(b + d)}$$

$$n = a + b + c + d = a + b + 0 + d = a + b + d \rightarrow a + b = n - d \text{ \& } b + d = n - a$$

$$A_{34}(j, k)^{(a,b,c=0,d)} = \frac{a[d + (n - a)]}{2(n - d)} + \frac{d[a + (n - d)]}{2(n - a)} < n$$

$$\text{máx. } A_{34}(j, k)^{(a,b,c,d)} = A_{34}(j, k)^{(a,b=c=0,d)} > A_{34}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{34}(j, k)^{(a,b,c=0,d)} - \\ \rightarrow A_{34}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA}$$

**Hawkins & Dotson** [0, 1]

$$A_{35}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{(a + b + c)} + \frac{d}{(d + b + c)} \right)$$

$$A_{35}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{35}(j, k)^{(a,b=0,c,d)} = \frac{1}{2} \left( \frac{a}{(a+0+c)} + \frac{d}{(d+0+c)} \right) = \frac{1}{2} \left( \frac{a}{a+c} + \frac{d}{d+c} \right)$$

$$= \frac{1}{2} [(expresión < 1) + (expresión < 1)] < 1$$

$$A_{35}(j, k)^{(a,b,c=0,d)} = \frac{1}{2} \left( \frac{a}{(a+b+0)} + \frac{d}{(d+b+0)} \right) = \frac{1}{2} \left( \frac{a}{a+b} + \frac{d}{d+b} \right)$$

$$= \frac{1}{2} [(expresión < 1) + (expresión < 1)] < 1$$

máx.  $A_{35}(j, k)^{(a,b,c,d)} = A_{35}(j, k)^{(a,b=c=0,d)} > A_{35}(j, k)^{(a,b=0,c,d)}$  y  $> A_{35}(j, k)^{(a,b,c=0,d)} \rightarrow A_{35}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Hurlbert\_1** [-1, 1]

$$A_{36}(j, k)^{(a,b,c,d)} = \pm \sqrt{\frac{\chi^2}{\chi_{máx.}^2}} = \pm \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{máx.}^2}}$$

$$= \pm \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{máx.}^2}}$$

$A_{36}(j, k)^{(a,b=c=0,d)} = 1$  (máximo) [con el signo de  $(ad-bc)$ ]

$$A_{36}(j, k)^{(a,b=0,c,d)} = + \sqrt{\frac{(a+0+c+d)(ad-0c)^2}{(a+0)(c+d)(a+c)(0+d) \chi_{máx.}^2}}$$

$$= + \sqrt{\frac{(a+c+d)(ad)^2}{a(c+d)(a+c)d \chi_{máx.}^2}} = + \sqrt{\frac{(a+c+d)(ad)^2}{ad(a+c)(c+d) \chi_{máx.}^2}} = + \sqrt{\frac{(a+c+d)ad}{(a+c)(c+d) \chi_{máx.}^2}}$$

Si  $\chi_{máx.}^2 = n = a+b+c+d = a+0+c+d = a+c+d \rightarrow$

$$A_{36}(j, k)^{(a,b=0,c,d)} = + \sqrt{\frac{(a+c+d)ad}{(a+c)(c+d)}} = + \sqrt{\frac{(a+c+d)ad}{(a+c)(c+d)(a+c+d)}}$$

$$= + \sqrt{\frac{ad}{(a+c)(c+d)}} = + \sqrt{\frac{ad}{ac+ad+c^2+cd}} = + \sqrt{\frac{ad}{ad+(ac+cd+c^2)}}$$

$$< 1$$

$$A_{36}(j, k)^{(a,b,c=0,d)} = + \sqrt{\frac{(a+b+0+d)(ad-b0)^2}{(a+b)(0+d)(a+0)(b+d) \chi_{máx.}^2}}$$

$$= + \sqrt{\frac{(a+b+d)(ad)^2}{(a+b)da(b+d) \chi_{máx.}^2}} = + \sqrt{\frac{(a+b+d)(ad)^2}{ad(a+b)(b+d) \chi_{máx.}^2}} = + \sqrt{\frac{(a+b+d)ad}{(a+b)(b+d) \chi_{máx.}^2}}$$

$$\text{Si } \chi_{\text{máx.}}^2 = n = a + b + c + d = a + b + 0 + d = a + b + d \rightarrow$$

$$\begin{aligned} A_{36}(j, k)^{(a,b,c=0,d)} &= + \sqrt{\frac{(a+b+d)ad}{(a+b)(b+d)}} = + \sqrt{\frac{(a+b+d)ad}{(a+b)(b+d)(a+b+d)}} \\ &= + \sqrt{\frac{ad}{(a+b)(b+d)}} = + \sqrt{\frac{ad}{ab+ad+b^2+bd}} = + \sqrt{\frac{ad}{ad+(ab+bd+b^2)}} \\ &< 1 \end{aligned}$$

$$\begin{aligned} \text{máx. } A_{36}(j, k)^{(a,b,c,d)} &= A_{36}(j, k)^{(a,b=c=0,d)} > A_{36}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{36}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{36}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Hurlbert\_2** [-1, 1]

$$\begin{aligned} A_{37}(j, k)^{(a,b,c,d)} &= \pm \sqrt{\frac{\chi^2 - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \pm \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \\ &= \pm \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \end{aligned}$$

$$A_{37}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo) [con el signo de } (ad - bc)]$$

$$\begin{aligned} A_{37}(j, k)^{(a,b=0,c,d)} &= + \sqrt{\frac{(a+0+c+d)(ad-0c)^2}{(a+0)(c+d)(a+c)(0+d)} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \\ &= + \sqrt{\frac{(a+c+d)(ad)^2}{a(c+d)(a+c)d} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = + \sqrt{\frac{(a+c+d)(ad)^2}{ad(c+d)(a+c)} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \\ &+ \sqrt{\frac{ad(a+c+d)}{(a+c)(c+d)} - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \end{aligned}$$

$$\text{Si } \chi_{\text{máx.}}^2 = n = a + b + c + d = a + 0 + c + d = a + c + d \text{ y } \chi_{\text{mín.}}^2 = 0 \rightarrow$$

$$\begin{aligned} A_{37}(j, k)^{(a,b=0,c,d)} &= + \sqrt{\frac{ad(a+c+d)}{ac+ad+c^2+cd} - 0}{(a+c+d) - 0}} = + \sqrt{\frac{ad(a+c+d)}{ad+(ac+cd+c^2)}} \\ &= + \sqrt{\frac{ad(a+c+d)}{ad+(ac+cd+c^2)} \frac{1}{(a+c+d)}} = + \sqrt{\frac{ad}{ad+(ac+cd+c^2)}} < 1 \end{aligned}$$

$$\begin{aligned}
A_{37}(j, k)^{(a,b,c=0,d)} &= + \sqrt{\frac{(a+b+0+d)(ad-b0)^2}{(a+b)(0+d)(a+0)(b+d)} - \chi_{\min}^2} \\
&= + \sqrt{\frac{(a+b+d)(ad)^2}{(a+b)da(b+d)} - \chi_{\min}^2} = + \sqrt{\frac{(a+b+d)(ad)^2}{ad(a+b)(b+d)} - \chi_{\min}^2} = \\
&+ \sqrt{\frac{ad(a+b+d)}{(a+b)(b+d)} - \chi_{\min}^2}
\end{aligned}$$

Si  $\chi_{\max}^2 = n = a + b + c + d = a + b + 0 + d = a + b + d$  y  $\chi_{\min}^2 = 0 \rightarrow$

$$\begin{aligned}
A_{37}(j, k)^{(a,b,c=0,d)} &= + \sqrt{\frac{ad(a+b+d)}{ab+ad+b^2+bd} - 0} = + \sqrt{\frac{ad(a+b+d)}{ad+(ab+bd+b^2)}} \\
&= + \sqrt{\frac{ad(a+b+d)}{ad+(ab+bd+b^2)} \frac{1}{(a+b+d)}} = + \sqrt{\frac{ad}{ad+(ab+bd+b^2)}} < 1
\end{aligned}$$

máx.  $A_{37}(j, k)^{(a,b,c,d)} = A_{37}(j, k)^{(a,b=c=0,d)} > A_{37}(j, k)^{(a,b=0,c,d)}$  y  $> A_{37}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{37}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Jaccard** [0, 1]

$$A_{38}(j, k)^{(a,b,c,d)} = \frac{a}{a+b+c}$$

$$A_{38}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{38}(j, k)^{(a,b=0,c,d)} = \frac{a}{a+0+c} = \frac{a}{a+c} < 1$$

$$A_{38}(j, k)^{(a,b,c=0,d)} = \frac{a}{a+b+0} = \frac{a}{a+b} < 1$$

máx.  $A_{38}(j, k)^{(a,b,c,d)} = A_{38}(j, k)^{(a,b=c=0,d)} > A_{38}(j, k)^{(a,b=0,c,d)}$  y  $> A_{38}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{38}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Johnson\_1** [0, 2]

$$A_{39}(j, k)^{(a,b,c,d)} = \frac{a}{a+b} + \frac{a}{a+c}$$

$$A_{39}(j, k)^{(a,b=c=0,d)} = 2 \text{ (máximo)}$$

$$A_{39}(j, k)^{(a,b=0,c,d)} = \frac{a}{a+0} + \frac{a}{a+c} = \frac{a}{a} + \frac{a}{a+c} = 1 + \frac{a}{a+c} = 1 + (\text{expresión} < 1) < 2$$

$$A_{39}(j, k)^{(a,b,c=0,d)} = \frac{a}{a+b} + \frac{a}{a+0} = \frac{a}{a+b} + \frac{a}{a} = \frac{a}{a+b} + 1 = 1 + (\text{expresión} < 1) < 2$$

$$\begin{aligned} \text{máx. } A_{39}(j, k)^{(a,b,c,d)} &= A_{39}(j, k)^{(a,b=c=0,d)} > A_{39}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{39}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{39}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Johnson\_2** [0, +∞]

$$A_{40}(j, k)^{(a,b,c,d)} = \frac{a}{2b}$$

$$A_{40}(j, k)^{(a,b=c=0,d)} \rightarrow +\infty \text{ (máximo)}$$

$$A_{40}(j, k)^{(a,b=0,c,d)} = \frac{a}{2 * 0} = \frac{a}{0} \rightarrow +\infty$$

$$A_{40}(j, k)^{(a,b,c=0,d)} = \frac{a}{2b} < +\infty$$

$$\begin{aligned} \text{máx. } A_{40}(j, k)^{(a,b,c,d)} &= A_{40}(j, k)^{(a,b=c=0,d)} = A_{40}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{40}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{40}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA (} b \\ &= 0) \text{ Y ABSOLUTA (} c = 0) \end{aligned}$$

**Kendall** [-1, 1]

$$A_{41}(j, k)^{(a,b,c,d)}$$

$$= \frac{2(ad - bc)}{\sqrt{(n^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)(n^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}} = \frac{2(ad - bc)}{\sqrt{([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

$$A_{41}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{41}(j, k)^{(a,b=0,c,d)}$$

$$\begin{aligned} &= \frac{2(ad - 0c)}{\sqrt{([a + 0 + c + d]^2 - a^2 - 0^2 - c^2 - d^2 + 2a0 + 2cd)([a + 0 + c + d]^2 - a^2 - 0^2 - c^2 - d^2 + 2ac + 2d0)}} \\ &= \frac{2ad}{\sqrt{([a + c + d]^2 - a^2 - c^2 - d^2 + 2cd)([a + c + d]^2 - a^2 - c^2 - d^2 + 2ac)}} = \frac{2ad}{\sqrt{(1) * (2)}} \end{aligned}$$

$$\begin{aligned} (1) &= [(a + c) + d]^2 - a^2 - c^2 - d^2 + 2cd \\ &= [(a + c)^2 + d^2 + 2d(a + c) - a^2 - c^2 - d^2 + 2cd] \\ &= a^2 + c^2 + 2ac + d^2 + 2ad + 2cd - a^2 - c^2 - d^2 + 2cd \\ &= 2ac + 2ad + 4cd \end{aligned}$$

$$\begin{aligned} (2) &= [(a + c) + d]^2 - a^2 - c^2 - d^2 + 2ac \\ &= [(a + c)^2 + d^2 + 2d(a + c) - a^2 - c^2 - d^2 + 2ac] \\ &= a^2 + c^2 + 2ac + d^2 + 2ad + 2cd - a^2 - c^2 - d^2 + 2ac \\ &= 4ac + 2ad + 2cd \end{aligned}$$

$$A_{41}(j, k)^{(a,b=0,c,d)} = \frac{2ad}{\sqrt{(1) * (2)}} = \frac{2ad}{\sqrt{(2ad + [2ac + 4cd]) * (2ad + [4ac + 2cd])}} \leq 1$$

$$\begin{aligned}
& A_{41}(j, k)^{(a,b,c=0,d)} \\
&= \frac{2(ad - b0)}{\sqrt{([a + b + 0 + d]^2 - a^2 - b^2 - 0^2 - d^2 + 2ab + 20d)([a + b + 0 + d]^2 - a^2 - b^2 - 0^2 - d^2 + 2a0 + 2bd)}} \\
&= \frac{2ad}{\sqrt{([a + b + d]^2 - a^2 - b^2 - d^2 + 2ab)([a + b + d]^2 - a^2 - b^2 - d^2 + 2bd)}} \\
&= \frac{2ad}{\sqrt{(3) * (4)}}
\end{aligned}$$

$$\begin{aligned}
(3) &= [(a + b) + d]^2 - a^2 - b^2 - d^2 + 2ab \\
&= [(a + b)^2 + d^2 + 2d(a + b) - a^2 - b^2 - d^2 + 2ab] \\
&= a^2 + b^2 + 2ab + d^2 + 2ad + 2bd - a^2 - b^2 - d^2 + 2ab \\
&= 4ab + 2ad + 2bd
\end{aligned}$$

$$\begin{aligned}
(4) &= [(a + b) + d]^2 - a^2 - b^2 - d^2 + 2bd \\
&= [(a + b)^2 + d^2 + 2d(a + b) - a^2 - b^2 - d^2 + 2bd] \\
&= a^2 + b^2 + 2ab + d^2 + 2ad + 2bd - a^2 - b^2 - d^2 + 2bd \\
&= 2ab + 2ad + 4bd
\end{aligned}$$

$$A_{41}(j, k)^{(a,b,c=0,d)} = \frac{2ad}{\sqrt{(3) * (4)}} = \frac{2ad}{\sqrt{(2ad + [4ab + 2bd]) * (2ad + [2ab + 4bd])}} \leq 1$$

*máx.  $A_{41}(j, k)^{(a,b,c,d)} = A_{41}(j, k)^{(a,b=c=0,d)} > A_{41}(j, k)^{(a,b=0,c,d)}$  y  $> A_{41}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{41}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA*

**Kuder & Richardson**  $[-\infty, 1]$

$$A_{42}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a + b)(c + d) + (a + c)(b + d) + 2(ad - bc)}$$

$$A_{42}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned}
A_{42}(j, k)^{(a,b=0,c,d)} &= \frac{4(ad - 0c)}{(a + 0)(c + d) + (a + c)(0 + d) + 2(ad - 0c)} \\
&= \frac{4ad}{4ad} \\
&= \frac{a(c + d) + (a + c)d + 2(ad)}{ac + ad + ad + cd + 2ad} \\
&= \frac{4ad}{4ad + (ac + cd)} < 1
\end{aligned}$$

$$\begin{aligned}
A_{42}(j, k)^{(a,b,c=0,d)} &= \frac{4(ad - b0)}{(a + b)(0 + d) + (a + 0)(b + d) + 2(ad - b0)} \\
&= \frac{4ad}{4ad} \\
&= \frac{(a + b)d + a(b + d) + 2(ad)}{ad + bd + ab + ad + 2ad} \\
&= \frac{4ad}{4ad + (bd + ab)} < 1
\end{aligned}$$

*máx.  $A_{42}(j, k)^{(a,b,c,d)} = A_{42}(j, k)^{(a,b=c=0,d)} > A_{42}(j, k)^{(a,b=0,c,d)}$  y  $> A_{42}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{42}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA*

**Kulczynski\_1**  $[0, +\infty]$

$$A_{43}(j, k)^{(a,b,c,d)} = \frac{a}{b+c}$$

$$A_{43}(j, k)^{(a,b=c=0,d)} \rightarrow +\infty \text{ (m\u00e1ximo)}$$

$$A_{43}(j, k)^{(a,b=0,c,d)} = \frac{a}{0+c} = \frac{a}{c} < +\infty$$

$$A_{43}(j, k)^{(a,b,c=0,d)} = \frac{a}{b+0} = \frac{a}{b} < +\infty$$

$$\text{m\u00e1x. } A_{43}(j, k)^{(a,b,c,d)} = A_{43}(j, k)^{(a,b=c=0,d)} > A_{43}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{43}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{43}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N ABSOLUTA

### **Kulczynski\_2** [0, 1]

$$A_{44}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{a+b} + \frac{a}{a+c} \right)$$

$$A_{44}(j, k)^{(a,b=c=0,d)} = 1 \text{ (m\u00e1ximo)}$$

$$A_{44}(j, k)^{(a,b=0,c,d)} = \frac{1}{2} \left( \frac{a}{a+0} + \frac{a}{a+c} \right) = \frac{1}{2} \left( \frac{a}{a} + \frac{a}{a+c} \right) = \frac{1}{2} \left( 1 + \frac{a}{a+c} \right)$$

$$= \frac{1}{2} (1 + [\text{expresi\u00f3n} < 1]) < 1$$

$$A_{44}(j, k)^{(a,b,c=0,d)} = \frac{1}{2} \left( \frac{a}{a+b} + \frac{a}{a+0} \right) = \frac{1}{2} \left( \frac{a}{a+b} + \frac{a}{a} \right) = \frac{1}{2} \left( \frac{a}{a+b} + 1 \right)$$

$$= \frac{1}{2} (1 + [\text{expresi\u00f3n} < 1]) < 1$$

$$\text{m\u00e1x. } A_{44}(j, k)^{(a,b,c,d)} = A_{44}(j, k)^{(a,b=c=0,d)} > A_{44}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{44}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{44}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N ABSOLUTA

### **Lamont & Grant** [0, 0.5]

$$A_{45}(j, k)^{(a,b,c,d)} = \frac{a}{2a+b+c}$$

$$A_{45}(j, k)^{(a,b=c=0,d)} = 0.5 \text{ (m\u00e1ximo)}$$

$$A_{45}(j, k)^{(a,b=0,c,d)} = \frac{a}{2a+0+c} = \frac{a}{2a+c} < 0.5$$

$$A_{45}(j, k)^{(a,b,c=0,d)} = \frac{a}{2a+b+0} = \frac{a}{2a+b} < 0.5$$

$$\text{m\u00e1x. } A_{45}(j, k)^{(a,b,c,d)} = A_{45}(j, k)^{(a,b=c=0,d)} > A_{45}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{45}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{45}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N ABSOLUTA

### **Legendre & Legendre** [0, 1]

$$A_{46}(j, k)^{(a,b,c,d)} = \frac{3a}{3a+b+c}$$

$$A_{46}(j, k)^{(a,b=c=0,d)} = 1 \text{ (m\u00e1ximo)}$$

$$A_{46}(j, k)^{(a,b=0,c,d)} = \frac{3a}{3a+0+c} = \frac{3a}{3a+c} < 1$$

$$A_{46}(j, k)^{(a,b,c=0,d)} = \frac{3a}{3a+b+0} = \frac{3a}{3a+b} < 1$$

$$\text{m\u00e1x. } A_{46}(j, k)^{(a,b,c,d)} = A_{46}(j, k)^{(a,b=c=0,d)} > A_{46}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{46}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{46}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N ABSOLUTA

**Loevinger**  $[-\infty, 1]$

$$A_{47}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{m\u00edn. } \{(a+b)(b+d), (a+c)(c+d)\}}$$

$$A_{47}(j, k)^{(a,b=c=0,d)} = 1 \text{ (m\u00e1ximo)}$$

$$A_{47}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{\text{m\u00edn. } \{(a+0)(0+d), (a+c)(c+d)\}} = \frac{ad}{\text{m\u00edn. } \{(ad), (a+c)(c+d)\}}$$

$$= \frac{ad}{ad} = 1$$

$$A_{47}(j, k)^{(a,b,c=0,d)} = \frac{ad - b0}{\text{m\u00edn. } \{(a+b)(b+d), (a+0)(0+d)\}} = \frac{ad}{\text{m\u00edn. } \{(a+b)(b+d), (ad)\}}$$

$$= \frac{ad}{ad} = 1$$

$$\text{m\u00e1x. } A_{47}(j, k)^{(a,b,c,d)} = A_{47}(j, k)^{(a,b=c=0,d)} = A_{47}(j, k)^{(a,b=0,c,d)} \text{ y } = A_{47}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{47}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N COMPLETA

**Maxwell & Pilliner**  $[-1, 1]$

$$A_{48}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(c+d) + (a+c)(b+d)}$$

$$A_{48}(j, k)^{(a,b=c=0,d)} = 1 \text{ (m\u00e1ximo)}$$

$$A_{48}(j, k)^{(a,b=0,c,d)} = \frac{2(ad - 0c)}{(a+0)(c+d) + (a+c)(0+d)} = \frac{2ad}{a(c+d) + (a+c)d}$$

$$= \frac{2ad}{\frac{2ad}{2ad} + \frac{2ad}{2ad}} = \frac{2ad}{2ad + ac + cd} = \frac{2ad}{2ad + (ac + cd)} < 1$$

$$A_{48}(j, k)^{(a,b,c=0,d)} = \frac{2(ad - b0)}{(a+b)(0+d) + (a+0)(b+d)} = \frac{2ad}{(a+b)d + a(b+d)}$$

$$= \frac{2ad}{\frac{2ad}{2ad} + \frac{2ad}{2ad}} = \frac{2ad}{2ad + ab + bd} = \frac{2ad}{2ad + (ab + bd)} < 1$$

$$\text{m\u00e1x. } A_{48}(j, k)^{(a,b,c,d)} = A_{48}(j, k)^{(a,b=c=0,d)} > A_{48}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{48}(j, k)^{(a,b,c=0,d)} \text{ --}$$

$\rightarrow A_{48}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N ABSOLUTA

**McConnaughey**  $[-1, 1]$

$$A_{49}(j, k)^{(a,b,c,d)} = \frac{a^2 - bc}{(a+b)(a+c)}$$

$$A_{49}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{49}(j, k)^{(a,b=0,c,d)} = \frac{a^2 - 0c}{(a+0)(a+c)} = \frac{a^2}{a(a+c)} = \frac{a^2}{a^2 + ac} < 1$$

$$A_{49}(j, k)^{(a,b,c=0,d)} = \frac{a^2 - b0}{(a+b)(a+0)} = \frac{a^2}{(a+b)a} = \frac{a^2}{a^2 + ab} < 1$$

máx.  $A_{49}(j, k)^{(a,b,c,d)} = A_{49}(j, k)^{(a,b=c=0,d)} > A_{49}(j, k)^{(a,b=0,c,d)}$  y  $> A_{49}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{49}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Michael** [-1, 1]

$$A_{50}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a+d)^2 + (b+c)^2}$$

$$A_{50}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo) Si } a = d$$

$$A_{50}(j, k)^{(a=d,b=0,c)} = \frac{4(dd - 0c)}{(d+d)^2 + (0+c)^2} = \frac{4d^2}{(2d)^2 + c^2} = \frac{4d^2}{4d^2 + c^2} < 1$$

$$A_{50}(j, k)^{(a=d,b,c=0)} = \frac{4(dd - b0)}{(d+d)^2 + (b+0)^2} = \frac{4d^2}{(2d)^2 + b^2} = \frac{4d^2}{4d^2 + b^2} < 1$$

máx.  $A_{50}(j, k)^{(a,b,c,d)} = A_{50}(j, k)^{(a=d,b=c=0)} > A_{50}(j, k)^{(a=d,b=0,c)}$  y  
 $> A_{50}(j, k)^{(a=d,b,c=0)}$  –  
 $\rightarrow A_{50}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA Si  $a = d$

**Mountford** [0, 2]

$$A_{51}(j, k)^{(a,b,c,d)} = \frac{a}{\frac{1}{2}(ab + ac) + bc}$$

$$A_{51}(j, k)^{(a,b=c=0,d)} \rightarrow \text{NO ES EL MÁXIMO}$$

**Ochiai** [0, 1]

$$A_{52}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

$$A_{52}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{52}(j, k)^{(a,b=0,c,d)} = \frac{a}{\sqrt{(a+0)(a+c)}} = \frac{a}{\sqrt{a(a+c)}} = \frac{a}{\sqrt{a^2 + ac}} < 1$$

$$A_{52}(j, k)^{(a,b,c=0,d)} = \frac{a}{\sqrt{(a+b)(a+0)}} = \frac{a}{\sqrt{(a+b)a}} = \frac{a}{\sqrt{a^2 + ab}} < 1$$

$$\begin{aligned} \text{máx. } A_{52}(j, k)^{(a,b,c,d)} &= A_{52}(j, k)^{(a,b=c=0,d)} > A_{52}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{52}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{52}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Pearson & Heron** [-1, 1]

$$A_{53}(j, k)^{(a,b,c,d)} = \cos \frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{53}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{53}(j, k)^{(a,b=0,c,d)} = \cos \frac{180\sqrt{0c}}{\sqrt{ad} + \sqrt{0c}} = \cos \frac{180 * 0}{\sqrt{ad}} = \cos(0) = 1$$

$$A_{53}(j, k)^{(a,b,c=0,d)} = \cos \frac{180\sqrt{b0}}{\sqrt{ad} + \sqrt{b0}} = \cos \frac{180 * 0}{\sqrt{ad}} = \cos(0) = 1$$

$$\begin{aligned} \text{máx. } A_{53}(j, k)^{(a,b,c,d)} &= A_{53}(j, k)^{(a,b=c=0,d)} = A_{53}(j, k)^{(a,b=0,c,d)} \text{ y } = A_{53}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{53}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA} \end{aligned}$$

**Pearson\_1** [0, n]

$$A_{54}(j, k)^{(a,b,c,d)} = \chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{54}(j, k)^{(a,b=c=0,d)} = n \text{ (máximo)}$$

$$\begin{aligned} A_{54}(j, k)^{(a,b=0,c,d)} &= \frac{(a + 0 + c + d)(ad - 0c)^2}{(a + 0)(c + d)(a + c)(0 + d)} = \frac{(a + c + d)(ad)^2}{a(c + d)(a + c)d} \\ &= \frac{(a + c + d)(ad)^2}{(a + c)(c + d)ad} = \frac{(a + c + d)ad}{(a + c)(c + d)} \end{aligned}$$

$$\begin{aligned} \text{Si } b = 0 \rightarrow n = a + b + c + d = a + c + d \rightarrow A_{54}(j, k)^{(a,b=0,c,d)} &= \frac{(a + c + d)ad}{(a + c)(c + d)} \\ &= \frac{n * ad}{(a + c)(c + d)} = n * \frac{ad}{(a + c)(c + d)} = n * (\text{expresión} < 1) < n \end{aligned}$$

$$\begin{aligned} A_{54}(j, k)^{(a,b,c=0,d)} &= \frac{(a + b + 0 + d)(ad - b0)^2}{(a + b)(0 + d)(a + 0)(b + d)} = \frac{(a + b + d)(ad)^2}{(a + b)da(b + d)} \\ &= \frac{(a + b + d)(ad)^2}{(a + b)(b + d)ad} = \frac{(a + b + d)ad}{(a + b)(b + d)} \end{aligned}$$

$$\begin{aligned} \text{Si } c = 0 \rightarrow n = a + b + c + d = a + b + d \rightarrow A_{54}(j, k)^{(a,b,c=0,d)} &= \frac{(a + b + d)ad}{(a + b)(b + d)} \\ &= \frac{n * ad}{(a + b)(b + d)} = n * \frac{ad}{(a + b)(b + d)} = n * (\text{expresión} < 1) < n \end{aligned}$$

$$\begin{aligned} \text{máx. } A_{54}(j, k)^{(a,b,c,d)} &= A_{54}(j, k)^{(a,b=c=0,d)} > A_{54}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{54}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{54}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Pearson\_2** [0, 1]

$$A_{55}(j, k)^{(a,b,c,d)} = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{(a+b+c+d) + \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

$A_{55}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Peirce\_1** [0, 1]

$$A_{56}(j, k)^{(a,b,c,d)} = \frac{ab + bc}{ab + 2bc + cd}$$

$A_{56}(j, k)^{(a,b=c=0,d)} \rightarrow$  NO ES EL MÁXIMO

**Peirce\_2** [-1, 1]

$$A_{57}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+c)(b+d)}$$

$A_{57}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{57}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{(a+c)(0+d)} = \frac{ad}{(a+c)d} = \frac{ad}{ad + (cd)} < 1$$

$$A_{57}(j, k)^{(a,b,c=0,d)} = \frac{ad - b0}{(a+0)(b+d)} = \frac{ad}{a(b+d)} = \frac{ad}{ab + ad} = \frac{ad}{ad + (ab)} < 1$$

$\text{máx. } A_{57}(j, k)^{(a,b,c,d)} = A_{57}(j, k)^{(a,b=c=0,d)} > A_{57}(j, k)^{(a,b=0,c,d)}$  y  $> A_{57}(j, k)^{(a,b,c=0,d)} \rightarrow$   
 $\rightarrow A_{57}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Peirce\_3** [-1, 1]

$$A_{58}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a+b)(c+d)}$$

$A_{58}(j, k)^{(a,b=c=0,d)} = 1$  (máximo)

$$A_{58}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{(a+0)(c+d)} = \frac{ad}{a(c+d)} = \frac{ad}{ac + ad} = \frac{ad}{ad + (ac)} < 1$$

$$A_{58}(j, k)^{(a,b,c=0,d)} = \frac{ad - b0}{(a+b)(0+d)} = \frac{ad}{(a+b)d} = \frac{ad}{ad + bd} = \frac{ad}{ad + (bd)} < 1$$

$\text{máx. } A_{58}(j, k)^{(a,b,c,d)} = A_{58}(j, k)^{(a,b=c=0,d)} > A_{58}(j, k)^{(a,b=0,c,d)}$  y  $> A_{58}(j, k)^{(a,b,c=0,d)} \rightarrow$   
 $\rightarrow A_{58}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Roger & Tanimoto** [0, 1]

$$A_{59}(j, k)^{(a,b,c,d)} = \frac{a + d}{a + 2(b + c) + d}$$

$$A_{59}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{59}(j, k)^{(a,b=0,c,d)} = \frac{a+d}{a+2(0+c)+d} = \frac{a+d}{a+2c+d} = \frac{a+d}{a+d+(2c)} < 1$$

$$A_{59}(j, k)^{(a,b,c=0,d)} = \frac{a+d}{a+2(b+0)+d} = \frac{a+d}{a+2b+d} = \frac{a+d}{a+d+(2b)} < 1$$

máx.  $A_{59}(j, k)^{(a,b,c,d)} = A_{59}(j, k)^{(a,b=c=0,d)} > A_{59}(j, k)^{(a,b=0,c,d)}$  y  $> A_{59}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{59}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Rogot & Goldberg** [0, 1]

$$A_{60}(j, k)^{(a,b,c,d)} = \frac{a}{2a+b+c} + \frac{d}{2d+b+c}$$

$$A_{60}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{60}(j, k)^{(a,b=0,c,d)} = \frac{a}{2a+0+c} + \frac{d}{2d+0+c} = \frac{a}{2a+c} + \frac{d}{2d+c} < 1$$

$$A_{60}(j, k)^{(a,b,c=0,d)} = \frac{a}{2a+b+0} + \frac{d}{2d+b+0} = \frac{a}{2a+b} + \frac{d}{2d+b} < 1$$

máx.  $A_{60}(j, k)^{(a,b,c,d)} = A_{60}(j, k)^{(a,b=c=0,d)} > A_{60}(j, k)^{(a,b=0,c,d)}$  y  $> A_{60}(j, k)^{(a,b,c=0,d)}$  –  
 →  $A_{60}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Russell & Rao** [0, 1]

$$A_{61}(j, k)^{(a,b,c,d)} = \frac{a}{n} = \frac{a}{a+b+c+d}$$

$$A_{61}(j, k)^{(a,b=c=0,d)} \rightarrow \text{NO ES EL MÁXIMO}$$

**Scott** [-1, 1]

$$A_{62}(j, k)^{(a,b,c,d)} = \frac{4(ad-bc) - (b-c)^2}{(2a+b+c)(2d+b+c)}$$

$$A_{62}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{62}(j, k)^{(a,b=0,c,d)} &= \frac{4(ad-0c) - (0-c)^2}{(2a+0+c)(2d+0+c)} = \frac{4ad-c^2}{(2a+c)(2d+c)} \\ &= \frac{4ad-c^2}{4ad+2ac+2cd+c^2} = \frac{4ad-c^2}{4ad+c^2+(2ac+2cd)} < 1 \end{aligned}$$

$$\begin{aligned} A_{62}(j, k)^{(a,b,c=0,d)} &= \frac{4(ad-b0) - (b-0)^2}{(2a+b+0)(2d+b+0)} = \frac{4ad-b^2}{(2a+b)(2d+b)} \\ &= \frac{4ad-b^2}{4ad+2ab+2bd+b^2} = \frac{4ad-b^2}{4ad+b^2+(2ab+2bd)} < 1 \end{aligned}$$

$$\begin{aligned} \text{máx. } A_{62}(j, k)^{(a,b,c,d)} &= A_{62}(j, k)^{(a,b=c=0,d)} > A_{62}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{62}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{62}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Simpson** [0, 1]

$$A_{63}(j, k)^{(a,b,c,d)} = \frac{a}{\text{mín.}\{(a+b), (a+c)\}}$$

$$A_{63}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{63}(j, k)^{(a,b=0,c,d)} = \frac{a}{\text{mín.}\{(a+0), (a+c)\}} = \frac{a}{\text{mín.}\{(a), (a+c)\}} = \frac{a}{a} = 1$$

$$A_{63}(j, k)^{(a,b,c=0,d)} = \frac{a}{\text{mín.}\{(a+b), (a+0)\}} = \frac{a}{\text{mín.}\{(a+b), (a)\}} = \frac{a}{a} = 1$$

$$\begin{aligned} \text{máx. } A_{63}(j, k)^{(a,b,c,d)} &= A_{63}(j, k)^{(a,b=c=0,d)} = A_{63}(j, k)^{(a,b=0,c,d)} \text{ y } = A_{63}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{63}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA} \end{aligned}$$

**Sokal & Michener** [0, 1]

$$A_{64}(j, k)^{(a,b,c,d)} = \frac{a+d}{n} = \frac{a+d}{a+b+c+d}$$

$$A_{64}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{64}(j, k)^{(a,b=0,c,d)} = \frac{a+d}{a+0+c+d} = \frac{a+d}{a+c+d} = \frac{a+d}{a+d+(c)} < 1$$

$$A_{64}(j, k)^{(a,b,c=0,d)} = \frac{a+d}{a+b+0+d} = \frac{a+d}{a+b+d} = \frac{a+d}{a+d+(b)} < 1$$

$$\begin{aligned} \text{máx. } A_{64}(j, k)^{(a,b,c,d)} &= A_{64}(j, k)^{(a,b=c=0,d)} > A_{64}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{64}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{64}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Sokal & Sneath\_1** [0, 1]

$$A_{65}(j, k)^{(a,b,c,d)} = \frac{a}{a+2(b+c)}$$

$$A_{65}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{65}(j, k)^{(a,b=0,c,d)} = \frac{a}{a+2(0+c)} = \frac{a}{a+2c} = \frac{a}{a+(2c)} < 1$$

$$A_{65}(j, k)^{(a,b,c=0,d)} = \frac{a}{a+2(b+0)} = \frac{a}{a+2b} = \frac{a}{a+(2b)} < 1$$

$$\begin{aligned} \text{máx. } A_{65}(j, k)^{(a,b,c,d)} &= A_{65}(j, k)^{(a,b=c=0,d)} > A_{65}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{65}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{65}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Sokal & Sneath\_2** [0, 1]

$$A_{66}(j, k)^{(a,b,c,d)} = \frac{1}{4} \left( \frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{b+d} \right)$$

$$A_{66}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{66}(j, k)^{(a,b=0,c,d)} &= \frac{1}{4} \left( \frac{a}{a+0} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{0+d} \right) = \frac{1}{4} \left( \frac{a}{a} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{d} \right) \\ &= \frac{1}{4} \left( 1 + \frac{a}{a+c} + \frac{d}{d+c} + 1 \right) = \frac{1}{4} \left( 2 + \frac{a}{a+c} + \frac{d}{d+c} \right) \\ &= \frac{1}{4} (2 + [\text{expresión} < 1] + [\text{expresión} < 1]) < 1 \end{aligned}$$

$$\begin{aligned} A_{66}(j, k)^{(a,b,c=0,d)} &= \frac{1}{4} \left( \frac{a}{a+b} + \frac{a}{a+0} + \frac{d}{0+d} + \frac{d}{b+d} \right) = \frac{1}{4} \left( \frac{a}{a+b} + \frac{a}{a} + \frac{d}{d} + \frac{d}{b+d} \right) \\ &= \frac{1}{4} \left( \frac{a}{a+b} + 1 + 1 + \frac{d}{d+b} \right) = \frac{1}{4} \left( 2 + \frac{a}{a+b} + \frac{d}{d+b} \right) \\ &= \frac{1}{4} (2 + [\text{expresión} < 1] + [\text{expresión} < 1]) < 1 \end{aligned}$$

$$\begin{aligned} \text{máx. } A_{66}(j, k)^{(a,b,c,d)} &= A_{66}(j, k)^{(a,b=c=0,d)} > A_{66}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{66}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{66}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

### Sokal & Sneath\_3 [0, +∞]

$$A_{67}(j, k)^{(a,b,c,d)} = \frac{a+d}{b+c}$$

$$A_{67}(j, k)^{(a,b=c=0,d)} \rightarrow +\infty \text{ (máximo)}$$

$$A_{67}(j, k)^{(a,b=0,c,d)} = \frac{a+d}{0+c} = \frac{a+d}{c} < +\infty$$

$$A_{67}(j, k)^{(a,b,c=0,d)} = \frac{a+d}{b+0} = \frac{a+d}{b} < +\infty$$

$$\begin{aligned} \text{máx. } A_{67}(j, k)^{(a,b,c,d)} &= A_{67}(j, k)^{(a,b=c=0,d)} > A_{67}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{67}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{67}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

### Sokal & Sneath\_4 [0, 1]

$$A_{68}(j, k)^{(a,b,c,d)} = \frac{a+d}{a + \frac{1}{2}(b+c) + d}$$

$$A_{68}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{68}(j, k)^{(a,b=0,c,d)} = \frac{a+d}{a + \frac{1}{2}(0+c) + d} = \frac{a+d}{a + \frac{1}{2}c + d} = \frac{a+d}{a+d + (\frac{1}{2}c)} < 1$$

$$A_{68}(j, k)^{(a,b,c=0,d)} = \frac{a+d}{a + \frac{1}{2}(b+0) + d} = \frac{a+d}{a + \frac{1}{2}b + d} = \frac{a+d}{a+d + (\frac{1}{2}b)} < 1$$

$\text{máx. } A_{68}(j, k)^{(a,b,c,d)} = A_{68}(j, k)^{(a,b=c=0,d)} > A_{68}(j, k)^{(a,b=0,c,d)}$  y  $> A_{68}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{68}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Sokal & Sneath\_5** [0, 1]

$$A_{69}(j, k)^{(a,b,c,d)} = \frac{ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$A_{69}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{69}(j, k)^{(a,b=0,c,d)} &= \frac{ad}{\sqrt{(a+0)(c+d)(a+c)(0+d)}} = \frac{ad}{\sqrt{a(c+d)(a+c)d}} \\ &= \frac{ad}{\sqrt{ad(a+c)(c+d)}} = \frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(a+c)(d+c)}} = \frac{\sqrt{ad}}{\sqrt{(a+c)(d+c)}} < 1 \end{aligned}$$

$$\begin{aligned} A_{69}(j, k)^{(a,b,c=0,d)} &= \frac{ad}{\sqrt{(a+b)(0+d)(a+0)(b+d)}} = \frac{ad}{\sqrt{(a+b)da(b+d)}} \\ &= \frac{ad}{\sqrt{ad(a+b)(b+d)}} = \frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(a+b)(d+b)}} = \frac{\sqrt{ad}}{\sqrt{(a+b)(d+b)}} < 1 \end{aligned}$$

$\text{máx. } A_{69}(j, k)^{(a,b,c,d)} = A_{69}(j, k)^{(a,b=c=0,d)} > A_{69}(j, k)^{(a,b=0,c,d)}$  y  $> A_{69}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{69}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Sorensen** [0, 1]

$$A_{70}(j, k)^{(a,b,c,d)} = \frac{2a}{2a+b+c}$$

$$A_{70}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{70}(j, k)^{(a,b=0,c,d)} = \frac{2a}{2a+0+c} = \frac{2a}{2a+c} < 1$$

$$A_{70}(j, k)^{(a,b,c=0,d)} = \frac{2a}{2a+b+0} = \frac{2a}{2a+b} < 1$$

$\text{máx. } A_{70}(j, k)^{(a,b,c,d)} = A_{70}(j, k)^{(a,b=c=0,d)} > A_{70}(j, k)^{(a,b=0,c,d)}$  y  $> A_{70}(j, k)^{(a,b,c=0,d)}$  –  
 $\rightarrow A_{70}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Sorgenfrei** [0, 1]

$$A_{71}(j, k)^{(a,b,c,d)} = \frac{a^2}{(a+b)(a+c)}$$

$$A_{71}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{71}(j, k)^{(a,b=0,c,d)} = \frac{a^2}{(a+0)(a+c)} = \frac{a^2}{a(a+c)} = \frac{a^2}{a^2+ac} = \frac{a^2}{a^2+(ac)} < 1$$

$$A_{71}(j, k)^{(a,b,c=0,d)} = \frac{a^2}{(a+b)(a+0)} = \frac{a^2}{(a+b)a} = \frac{a^2}{a^2+ab} = \frac{a^2}{a^2+(ab)} < 1$$

$$\begin{aligned} \text{máx. } A_{71}(j, k)^{(a,b,c,d)} &= A_{71}(j, k)^{(a,b=c=0,d)} > A_{71}(j, k)^{(a,b=0,c,d)} \text{ y } > A_{71}(j, k)^{(a,b,c=0,d)} - \\ &\rightarrow A_{71}(j, k) \text{ ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA} \end{aligned}$$

**Stiles**  $[-\infty, +\infty]$

$$\begin{aligned} A_{72}(j, k)^{(a,b,c,d)} &= \log \frac{n(|ad - bc| - \frac{n}{2})^2}{(a+b)(a+c)(b+d)(c+d)} \\ &= \log \frac{(a+b+c+d)(|ad - bc| - \frac{a+b+c+d}{2})^2}{(a+b)(a+c)(b+d)(c+d)} \end{aligned}$$

$$A_{72}(j, k)^{(a,b=c=0,d)} \rightarrow \text{NO ES EL MÁXIMO}$$

**Tarwid**  $[-1, 1]$

$$A_{73}(j, k)^{(a,b,c,d)} = \frac{na - (a+b)(a+c)}{na + (a+b)(a+c)} = \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d)a + (a+b)(a+c)}$$

$$A_{73}(j, k)^{(a,b=c=0,d)} \rightarrow \text{NO ES EL MÁXIMO}$$

**Tschuproff**  $[0, 1]$

$$\begin{aligned} A_{74}(j, k)^{(a,b,c,d)} &= \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{\frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{a+b+c+d}} \\ &= \sqrt{\frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}} \end{aligned}$$

$$A_{74}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned} A_{74}(j, k)^{(a,b=0,c,d)} &= \sqrt{\frac{(a+0+c+d)(ad - 0c)^2}{(a+0)(c+d)(a+c)(0+d)}} \\ &= \sqrt{\frac{(a+c+d)(ad)^2}{a(c+d)(a+c)d}} = \sqrt{\frac{(a+c+d)(ad)^2}{ad(c+d)(a+c)}} = \sqrt{\frac{(a+c+d)ad}{(c+d)(a+c)}} \\ &= \sqrt{\frac{(a+c+d)ad}{(a+c)(d+c)} \frac{1}{(a+c+d)}} = \sqrt{\frac{ad}{(a+c)(d+c)}} < 1 \end{aligned}$$

$$\begin{aligned}
A_{74}(j, k)^{(a,b,c=0,d)} &= \sqrt{\frac{(a+b+0+d)(ad-b0)^2}{(a+b)(0+d)(a+0)(b+d)}} \\
&= \sqrt{\frac{(a+b+d)(ad)^2}{(a+b)da(b+d)}} = \sqrt{\frac{(a+b+d)(ad)^2}{ad(a+b)(b+d)}} = \sqrt{\frac{(a+b+d)ad}{(a+b)(d+b)}} \\
&= \sqrt{\frac{(a+b+d)ad}{(a+b)(d+b)} \cdot \frac{1}{a+b+d}} = \sqrt{\frac{ad}{(a+b)(d+b)}} < 1
\end{aligned}$$

máx.  $A_{74}(j, k)^{(a,b,c,d)} = A_{74}(j, k)^{(a,b=c=0,d)} > A_{74}(j, k)^{(a,b=0,c,d)}$  y  $> A_{74}(j, k)^{(a,b,c=0,d)}$  –  
→  $A_{74}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Yule \_1** [-1, 1]

$$A_{75}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \phi$$

$$A_{75}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$\begin{aligned}
A_{75}(j, k)^{(a,b=0,c,d)} &= \frac{ad - 0c}{\sqrt{(a+0)(c+d)(a+c)(0+d)}} = \frac{ad}{\sqrt{a(c+d)(a+c)d}} \\
&= \frac{ad}{\sqrt{ad(a+c)(d+c)}} = \frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(a+c)(d+c)}} = \frac{\sqrt{ad}}{\sqrt{(a+c)(d+c)}} < 1
\end{aligned}$$

$$\begin{aligned}
A_{75}(j, k)^{(a,b,c=0,d)} &= \frac{ad - b0}{\sqrt{(a+b)(0+d)(a+0)(b+d)}} = \frac{ad}{\sqrt{(a+b)da(b+d)}} \\
&= \frac{ad}{\sqrt{ad(a+b)(b+d)}} = \frac{\sqrt{ad}\sqrt{ad}}{\sqrt{ad}\sqrt{(a+b)(d+b)}} = \frac{\sqrt{ad}}{\sqrt{(a+b)(d+b)}} < 1
\end{aligned}$$

máx.  $A_{75}(j, k)^{(a,b,c,d)} = A_{75}(j, k)^{(a,b=c=0,d)} > A_{75}(j, k)^{(a,b=0,c,d)}$  y  $> A_{75}(j, k)^{(a,b,c=0,d)}$  –  
→  $A_{75}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN ABSOLUTA

**Yule \_2** [-1, 1]

$$A_{76}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{ad + bc}$$

$$A_{76}(j, k)^{(a,b=c=0,d)} = 1 \text{ (máximo)}$$

$$A_{76}(j, k)^{(a,b=0,c,d)} = \frac{ad - 0c}{ad + 0c} = \frac{ad}{ad} = 1$$

$$A_{76}(j, k)^{(a,b,c=0,d)} = \frac{ad - b0}{ad + b0} = \frac{ad}{ad} = 1$$

máx.  $A_{76}(j, k)^{(a,b,c,d)} = A_{76}(j, k)^{(a,b=c=0,d)} = A_{76}(j, k)^{(a,b=0,c,d)}$  y  $= A_{76}(j, k)^{(a,b,c=0,d)}$  –  
→  $A_{76}(j, k)$  ES UN COEFICIENTE DE ASOCIACIÓN COMPLETA

**Yule \_3** [-1, 1]

$$A_{77}(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{77}(j, k)^{(a,b=c=0,d)} = 1 \text{ (m\u00e1ximo)}$$

$$A_{77}(j, k)^{(a,b=0,c,d)} = \frac{\sqrt{ad} - \sqrt{0c}}{\sqrt{ad} + \sqrt{0c}} = \frac{\sqrt{ad}}{\sqrt{ad}} = 1$$

$$A_{77}(j, k)^{(a,b,c=0,d)} = \frac{\sqrt{ad} - \sqrt{b0}}{\sqrt{ad} + \sqrt{b0}} = \frac{\sqrt{ad}}{\sqrt{ad}} = 1$$

m\u00e1x.  $A_{77}(j, k)^{(a,b,c,d)} = A_{77}(j, k)^{(a,b=c=0,d)} = A_{77}(j, k)^{(a,b=0,c,d)}$  y  $= A_{77}(j, k)^{(a,b,c=0,d)}$  –  
→  $A_{77}(j, k)$  ES UN COEFICIENTE DE ASOCIACI\u00d3N COMPLETA



## ANEXO 4

Propiedad 5: Asociación nula cuando el valor de  $a=0$

$$\text{mín. } A_i(j, k)^{(a,b,c,d)} = A_i(j, k)^{(a=0,b,c,d)}$$

**Anónimo\_1** [-1, 1]

$$A_1(j, k)^{(a,b,c,d)} = \frac{2a - (b + c)}{2a + (b + c)}$$

$$A_1(j, k)^{(a=0,b,c,d)} = \frac{2 * 0 - (b + c)}{2 * 0 + (b + c)} = \frac{-(b + c)}{(b + c)} = -1 \text{ (mínimo)}$$

**Anónimo\_2** [0, 1]

$$A_2(j, k)^{(a,b,c,d)} = \frac{2d}{2d + b + c}$$

$$A_2(j, k)^{(a=0,b,c,d)} = \frac{2d}{2d + (b + c)} < 1 \text{ NO ES EL MÍNIMO}$$

**Anónimo\_3** [-1, 1]

$$A_3(j, k)^{(a,b,c,d)} = \frac{2d - b - c}{2d + b + c}$$

$$A_3(j, k)^{(a=0,b,c,d)} = \frac{2d - b - c}{2d + b + c} = \frac{2d - (b + c)}{2d + (b + c)} < 1 \text{ NO ES EL MÍNIMO}$$

**Anónimo\_4** [0, 1]

$$A_4(j, k)^{(a,b,c,d)} = \frac{4ad}{4ad + (a + d)(b + c)}$$

$$A_4(j, k)^{(a=0,b,c,d)} = \frac{4 * 0 * d}{4 * 0 * d + (0 + d)(b + c)} = \frac{0}{0 + d(b + c)} = \frac{0}{d(b + c)} = 0 \text{ (mínimo)}$$

**Anónimo\_5** [-1, +∞]

$$A_5(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a + b)(a + c), (c + d)(b + d)\}}$$

$$\begin{aligned}
A_5(j, k)^{(a=0, b, c, d)} &= \frac{0d - bc}{\min\{(0 + b)(0 + c), (c + d)(b + d)\}} = \frac{-bc}{\min\{bc, (c + d)(b + d)\}} \\
&= \frac{-bc}{\min\{bc, (cb + cd + db + d^2)\}} = \frac{-bc}{\min\{bc, (bc + bd + cd + d^2)\}} \\
&= \frac{-bc}{bc + (bd + cd + d^2)} > -1 \text{ NO ES EL MÍNIMO}
\end{aligned}$$

**Baroni-Urbani & Buser\_1** [0, 1]

$$A_6(j, k)^{(a, b, c, d)} = \frac{\sqrt{ad} + a}{\sqrt{ad} + a + b + c}$$

$$A_6(j, k)^{(a=0, b, c, d)} = \frac{\sqrt{0d} + 0}{\sqrt{0d} + 0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Baroni-Urbani & Buser\_2** [-1, 1]

$$A_7(j, k)^{(a, b, c, d)} = \frac{\sqrt{ad} + a - b - c}{\sqrt{ad} + a + b + c}$$

$$A_7(j, k)^{(a=0, b, c, d)} = \frac{\sqrt{0d} + 0 - b - c}{\sqrt{0d} + 0 + b + c} = \frac{-b - c}{b + c} = \frac{-(b + c)}{b + c} = -1 \text{ (mínimo)}$$

**Braun-Blanquet** [0, 1]

$$A_8(j, k)^{(a, b, c, d)} = \frac{a}{\max\{(a + b), (a + c)\}}$$

$$A_8(j, k)^{(a=0, b, c, d)} = \frac{0}{\max\{(0 + b), (0 + c)\}} = \frac{0}{\max\{(b), (c)\}} = 0 \text{ (mínimo)}$$

**Clement** [0, n]

$$A_9(j, k)^{(a, b, c, d)} = \frac{a(c + d)}{(a + b)} + \frac{d(a + b)}{(c + d)}$$

$$A_9(j, k)^{(a=0, b, c, d)} = \frac{0(c + d)}{(0 + b)} + \frac{d(0 + b)}{(c + d)} = \frac{0}{b} + \frac{d(b)}{c + d} = \frac{db}{c + d} \text{ NO ES EL MÍNIMO}$$

**Clifford & Stephenson** [0, 1]

$$A_{10}(j, k)^{(a, b, c, d)} = \frac{\chi^2}{n} = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$A_{10}(j, k)^{(a=0, b, c, d)} = \frac{(0d - bc)^2}{(0 + b)(c + d)(0 + c)(b + d)} = \frac{(-bc)^2}{b(c + d)c(b + b)} = \frac{(bc)^2}{bc(c + d)(b + d)}$$

$$= \frac{bc}{(c + d)(b + d)} = \text{NO ES EL MÍNIMO}$$

**Cohen** [-1, 1]

$$A_{11}(j, k)^{(a, b, c, d)} = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}$$

$$A_{11}(j, k)^{(a=0, b, c, d)} = \frac{2(0d - bc)}{(0 + b)(b + d) + (0 + c)(c + d)} = \frac{-2bc}{b(b + d) + c(c + d)}$$

$$= \frac{-2bc}{b^2 + bd + c^2 + cd} \quad \text{NO ES EL MÍNIMO}$$

**Cole\_1** [-1, 1]

$$A_{12}(j, k)^{(a, b, c, d)} = \frac{\sqrt{2}(ad - bc)}{\sqrt{(ad - bc)^2 + (a + b)(c + d)(a + c)(b + d)}}$$

$$A_{12}(j, k)^{(a=0, b, c, d)} = \frac{\sqrt{2}(0d - bc)}{\sqrt{(0d - bc)^2 + (0 + b)(c + d)(0 + c)(b + d)}}$$

$$= \frac{-\sqrt{2}bc}{\sqrt{(-bc)^2 + b(c + d)c(b + d)}} = \frac{-\sqrt{2}bc}{\sqrt{(bc)^2 + bc(c + d)(b + d)}}$$

$$= \frac{-\sqrt{2}bc}{\sqrt{(bc)[(bc) + (c + d)(b + d)]}} = \frac{-\sqrt{2}\sqrt{bc}\sqrt{bc}}{\sqrt{bc}\sqrt{(bc) + (c + d)(b + d)}}$$

$$= \frac{-\sqrt{2}\sqrt{bc}}{\sqrt{(bc) + (c + d)(b + d)}} \quad \text{NO ES EL MÍNIMO}$$

**Cole\_2** [0, 1]

$$A_{13}(j, k)^{(a, b, c, d)} = \frac{\phi}{\phi_{\text{máx.}}} = \frac{\frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}}{\phi_{\text{máx.}}}$$

$$A_{13}(j, k)^{(a=0, b, c, d)} = \frac{\frac{0d - bc}{\sqrt{(0 + b)(c + d)(0 + c)(b + d)}}}{\phi_{\text{máx.}}} = \frac{\frac{-bc}{\sqrt{bc(c + d)(b + d)}}}{\phi_{\text{máx.}}}$$

$$= \frac{\frac{-\sqrt{bc}\sqrt{bc}}{\sqrt{bc}\sqrt{(c + d)(b + d)}}}{\phi_{\text{máx.}}} = \frac{\frac{-\sqrt{bc}}{\sqrt{(c + d)(b + d)}}}{\phi_{\text{máx.}}}$$

$$\text{Si } \phi_{\text{máx.}} = 1 \rightarrow A_{13}(j, k)^{(a=0, b, c, d)} = \frac{-\sqrt{bc}}{\sqrt{(c + d)(b + d)}} \quad \text{NO ES EL MÍNIMO}$$

**Cole\_3**  $[-\infty, 1]$

$$A_{14}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + b)(b + d)}$$

$$A_{14}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{(0 + b)(b + d)} = \frac{-bc}{b(b + d)} = \frac{-c}{b + d} \text{ NO ES EL MÍNIMO}$$

**Cole\_4**  $[-\infty, 1]$

$$A_{15}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + c)(c + d)}$$

$$A_{15}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{(0 + c)(c + d)} = \frac{-bc}{c(c + d)} = \frac{-b}{c + d} \text{ NO ES EL MÍNIMO}$$

**Dennis**  $[-\infty, +\infty]$

$$A_{16}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a + b + c + d)(a + b)(a + c)}}$$

$$\begin{aligned} A_{16}(j, k)^{(a=0,b,c,d)} &= \frac{0d - bc}{\sqrt{(0 + b + c + d)(0 + b)(0 + c)}} = \frac{-bc}{\sqrt{(b + c + d)bc}} = \frac{-\sqrt{bc}\sqrt{bc}}{\sqrt{bc}\sqrt{(b + c + d)}} \\ &= \frac{-\sqrt{bc}}{\sqrt{(b + c + d)}} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Dice\_1**  $[0, 1]$

$$A_{17}(j, k)^{(a,b,c,d)} = \frac{a}{a + b}$$

$$A_{17}(j, k)^{(a=0,b,c,d)} = \frac{0}{0 + b} = \frac{0}{b} = 0 \text{ (mínimo)}$$

**Dice\_2**  $[0, 1]$

$$A_{18}(j, k)^{(a,b,c,d)} = \frac{a}{a + c}$$

$$A_{18}(j, k)^{(a=0,b,c,d)} = \frac{0}{0 + c} = \frac{0}{c} = 0 \text{ (mínimo)}$$

**Dice\_3**  $[0, 2]$

$$A_{19}(j, k)^{(a,b,c,d)} = \frac{2a}{(a + b)(a + c)}$$

$$A_{19}(j, k)^{(a=0,b,c,d)} = \frac{2 \times 0}{(0 + b)(0 + c)} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Digby** [-1, 1]

$$A_{20}(j, k)^{(a,b,c,d)} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}}$$

$$A_{20}(j, k)^{(a=0,b,c,d)} = \frac{(0d)^{3/4} - (bc)^{3/4}}{(0d)^{3/4} + (bc)^{3/4}} = \frac{-(bc)^{3/4}}{(bc)^{3/4}} = -1 \text{ (mínimo)}$$

**Doolittle** [0, 1]

$$A_{21}(j, k)^{(a,b,c,d)} = \Phi^2 = \frac{(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$\begin{aligned} A_{21}(j, k)^{(a=0,b,c,d)} &= \frac{(0d - bc)^2}{(0 + b)((c + d)(0 + c)(b + d))} = \frac{(-bc)^2}{b(c + d)c(b + d)} = \frac{(bc)^2}{bc(c + d)(b + d)} \\ &= \frac{bc}{(c + d)(b + d)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Eyraud** [-1, 0]

$$A_{22}(j, k)^{(a,b,c,d)} = \frac{a - (a + b)(a + c)}{(a + b)(c + d)(a + c)(b + d)}$$

$$\begin{aligned} A_{22}(j, k)^{(a=0,b,c,d)} &= \frac{0 - (0 + b)(0 + c)}{(0 + b)(c + d)(0 + c)(b + d)} = \frac{-bc}{b(c + d)c(b + d)} = \frac{-bc}{bc(c + d)(b + d)} \\ &= \frac{-1}{(c + d)(b + d)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Fager & McGowan** [-1, 1]

$$A_{23}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a + b)(a + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(a + b), (a + c)\}}}$$

$$A_{23}(j, k)^{(a=0,b,c,d)} = \frac{0}{\sqrt{(0 + b)(0 + c)}} - \frac{1}{2\sqrt{\text{máx.}\{(0 + b), (0 + c)\}}} = -\frac{1}{2\sqrt{\text{máx.}\{(b), (c)\}}}$$

$$\text{Si } b > c \rightarrow A_{23}(j, k)^{(a=0,b,c,d)} = -\frac{1}{2\sqrt{b}} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b \leq c \rightarrow A_{23}(j, k)^{(a=0,b,c,d)} = -\frac{1}{2\sqrt{c}} \text{ NO ES EL MÍNIMO}$$

**Faith** [0, 1]

$$A_{24}(j, k)^{(a,b,c,d)} = \frac{a + \frac{d}{2}}{n} = \frac{a + \frac{d}{2}}{a + b + c + d}$$

$$A_{24}(j, k)^{(a=0, b, c, d)} = \frac{0 + \frac{d}{2}}{0 + b + c + d} = \frac{\frac{d}{2}}{b + c + d} = \frac{d}{2} \frac{1}{(b + c + d)}$$

$$= \frac{d}{2(b + c + d)} \text{ NO ES EL MÍNIMO}$$

**Fleiss**  $[-\infty, 1]$

$$A_{25}(j, k)^{(a, b, c, d)} = \frac{(ad - bc)[(a + b)(b + d) + (a + c)(c + d)]}{2(a + b)(a + c)(b + d)(c + d)}$$

$$A_{25}(j, k)^{(a=0, b, c, d)} = \frac{(0d - bc)[(0 + b)(b + d) + (0 + c)(c + d)]}{2(0 + b)(0 + c)(b + d)(c + d)}$$

$$= \frac{(-bc)[b(b + d) + c(c + d)]}{2bc(b + d)(c + d)} = \frac{(-bc)[b(b + d) + c(c + d)]}{2bc(b + d)(c + d)}$$

$$= \frac{-[b(b + d) + c(c + d)]}{2(b + d)(c + d)} \text{ NO ES EL MÍNIMO}$$

**Forbes\_1**  $[0, +\infty]$

$$A_{26}(j, k)^{(a, b, c, d)} = \frac{n * a}{(a + b)(a + c)} = \frac{(a + b + c + d)a}{(a + b)(a + c)}$$

$$A_{26}(j, k)^{(a=0, b, c, d)} = \frac{(0 + b + c + d)0}{(0 + b)(0 + c)} = \frac{(b + c + d) * 0}{bc} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Forbes\_2**  $[-1, 1]$

$$A_{27}(j, k)^{(a, b, c, d)} = \frac{n * a - (a + b)(a + c)}{n * \text{máx.}\{(a + b), (a + c)\} - (a + b)(a + c)}$$

$$= \frac{(a + b + c + d)a - (a + b)(a + c)}{(a + b + c + d) * \text{máx.}\{(a + b), (a + c)\} - (a + b)(a + c)}$$

$$A_{27}(j, k)^{(a=0, b, c, d)} = \frac{(0 + b + c + d)0 - (0 + b)(0 + c)}{(0 + b + c + d) * \text{máx.}\{(0 + b), (0 + c)\} - (0 + b)(0 + c)}$$

$$= \frac{(b + c + d) * 0 - (bc)}{(b + c + d) * \text{máx.}\{(b), (c)\} - bc} = \frac{-bc}{(b + c + d) * \text{máx.}\{(b), (c)\} - bc}$$

Si  $b > c \rightarrow A_{27}(j, k)^{(a=0, b, c, d)} = \frac{-bc}{(b + c + d)b - bc} = \frac{b(-c)}{b[(b + c + d) - c]}$

$$= \frac{-c}{(b + c + d) - c} \text{ NO ES EL MÍNIMO}$$

Si  $b < c \rightarrow A_{27}(j, k)^{(a=0, b, c, d)} = \frac{-bc}{(b + c + d)c - bc} = \frac{c(-b)}{c[(b + c + d) - b]}$

$$= \frac{-b}{(b + c + d) - b} \text{ NO ES EL MÍNIMO}$$

$$\begin{aligned} \text{Si } b = c \rightarrow A_{27}(j, k)^{(a=0, b, c, d)} &= \frac{-cc}{(c+c+d)c-cc} = \frac{c(-c)}{c[(2c+d)-c]} = \frac{-c}{2c+d-c} \\ &= \frac{-c}{c+d} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Forbes\_3**  $[-1, +\infty]$

$$A_{28}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a+b)(a+c)}$$

$$A_{28}(j, k)^{(a=0, b, c, d)} = \frac{0d - bc}{(0+b)(0+c)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Fossum**  $[0, +\infty]$

$$A_{29}(j, k)^{(a, b, c, d)} = \frac{n\left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)} = \frac{(a+b+c+d)\left(a - \frac{1}{2}\right)^2}{(a+b)(a+c)}$$

$$\begin{aligned} A_{29}(j, k)^{(a=0, b, c, d)} &= \frac{(0+b+c+d)\left(0 - \frac{1}{2}\right)^2}{(0+b)(0+c)} = \frac{(b+c+d)\frac{1}{4}}{bc} \\ &= \frac{(b+c+d)}{4bc} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Gilbert & Wells**  $[-\infty, +\infty]$

$$\begin{aligned} A_{30}(j, k)^{(a, b, c, d)} &= \log(a) - \log(n) - \log\left(\frac{a+b}{n}\right) - \log\left(\frac{a+c}{n}\right) \\ &= \log(a) - \log(a+b+c+d) - \log\left(\frac{a+b}{a+b+c+d}\right) - \log\left(\frac{a+c}{a+b+c+d}\right) \end{aligned}$$

$$\begin{aligned} A_{30}(j, k)^{(a=0, b, c, d)} &= \log(0) - \log(0+b+c+d) - \log\left(\frac{0+b}{0+b+c+d}\right) - \log\left(\frac{0+c}{0+b+c+d}\right) \\ &= \log(0) - \log(b+c+d) - \log\left(\frac{b}{b+c+d}\right) - \log\left(\frac{c}{b+c+d}\right) = \\ &\rightarrow -\infty \text{ (mínimo)} \end{aligned}$$

**Goodall**  $[0, 1]$

$$A_{31}(j, k)^{(a, b, c, d)} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{n}} = \frac{2}{\pi} \arcsin \sqrt{\frac{a+d}{a+b+c+d}}$$

$$\begin{aligned} A_{31}(j, k)^{(a=0, b, c, d)} &= \frac{2}{\pi} \arcsin \sqrt{\frac{0+d}{0+b+c+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{d}{b+c+d}} = \frac{2}{\pi} \arcsin \sqrt{\frac{d}{d+(b+c)}} \\ &= \frac{2}{\pi} \arcsin \sqrt{(\text{expresión} < 1)} = \frac{2}{\pi} \left(\text{expresión} < \frac{\pi}{2}\right) > 0 \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Goodman & Kruskal** [-1, 1]

$$A_{32}(j, k)^{(a,b,c,d)} = \frac{2\text{mín.}\{(a), (d)\} - b - c}{2\text{mín.}\{(a), (d)\} + b + c}$$

$$A_{32}(j, k)^{(a=0,b,c,d)} = \frac{2\text{mín.}\{(0), (d)\} - b - c}{2\text{mín.}\{(0), (d)\} + b + c} = \frac{2 * 0 - b - c}{2 * 0 + b + c} = \frac{-(b + c)}{b + c} = -1 \text{ (mínimo)}$$

**Hamann** [-1, 1]

$$A_{33}(j, k)^{(a,b,c,d)} = \frac{a + d - b - c}{n} = \frac{a + d - b - c}{a + d + b + c}$$

$$A_{33}(j, k)^{(a=0,b,c,d)} = \frac{0 + d - b - c}{0 + d + b + c} = \frac{d - (b + c)}{d + (b + c)} \text{ NO ES EL MÍNIMO}$$

**Harris & Lahey** [0, n]

$$A_{34}(j, k)^{(a,b,c,d)} = \frac{a(2d + b + c)}{2(a + b + c)} + \frac{d(2a + b + c)}{2(d + b + c)}$$

$$\begin{aligned} A_{34}(j, k)^{(a=0,b,c,d)} &= \frac{0(2d + b + c)}{2(0 + b + c)} + \frac{d(2 * 0 + b + c)}{2(d + b + c)} = \frac{0}{2(b + c)} + \frac{d(b + c)}{2(d + b + c)} \\ &= 0 + \frac{d(b + c)}{2(d + b + c)} = \frac{d(b + c)}{2(d + b + c)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Hawkins & Dotson** [0, 1]

$$A_{35}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{(a + b + c)} + \frac{d}{(d + b + c)} \right)$$

$$\begin{aligned} A_{35}(j, k)^{(a=0,b,c,d)} &= \frac{1}{2} \left( \frac{0}{(0 + b + c)} + \frac{d}{(d + b + c)} \right) = \frac{1}{2} \left( \frac{0}{b + c} + \frac{d}{d + b + c} \right) \\ &= \frac{1}{2} \left( \frac{d}{(d + (b + c))} \right) = \frac{1}{2} (\text{expresión} < 1) > 0 \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Hurlbert\_1** [-1, 1]

$$\begin{aligned} A_{36}(j, k)^{(a,b,c,d)} &= \pm \sqrt{\frac{\chi^2}{\chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2}} \\ &= \pm \sqrt{\frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d) \chi_{\text{máx.}}^2}} \end{aligned}$$

$$\begin{aligned}
A_{36}(j, k)^{(a=0,b,c,d)} &= \pm \sqrt{\frac{(0+b+c+d)(0d-bc)^2}{(0+b)(c+d)(0+c)(b+d) \chi_{\text{máx.}}^2}} \\
&= \pm \sqrt{\frac{(b+c+d)(bc)^2}{b(c+d)c(b+d) \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{bc^2(b+c+d)}{bc(c+d)(b+d) \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{bc(b+c+d)}{(c+d)(b+d) \chi_{\text{máx.}}^2}}
\end{aligned}$$

Si  $\chi_{\text{máx.}}^2 = n = a + b + c + d = 0 + b + c + d = b + c + d \rightarrow A_{36}(j, k)^{(a=0,b,c,d)}$

$$\begin{aligned}
&= \pm \sqrt{\frac{bc(b+c+d)}{(c+d)(b+d) \chi_{\text{máx.}}^2}} = \pm \sqrt{\frac{bc(b+c+d)}{(c+d)(b+d)} \frac{1}{b+c+d}} \\
&= \pm \sqrt{\frac{bc}{(c+d)(b+d)}} \text{ NO ES EL MÍNIMO}
\end{aligned}$$

## Hurlbert\_2 [-1, 1]

$$\begin{aligned}
A_{37}(j, k)^{(a,b,c,d)} &= \pm \sqrt{\frac{\chi^2 - \chi_{\text{mín.}}^2}{\chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \pm \sqrt{\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \\
&= \pm \sqrt{\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}}
\end{aligned}$$

$$\begin{aligned}
A_{37}(j, k)^{(a=0,b,c,d)} &= \pm \sqrt{\frac{(0+b+c+d)(0d-bc)^2}{(0+b)(c+d)(0+c)(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \\
&= \pm \sqrt{\frac{(b+c+d)(-bc)^2}{b(c+d)c(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} = \pm \sqrt{\frac{(b+c+d)(bc)^2}{bc(c+d)(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}} \\
&= \pm \sqrt{\frac{(b+c+d)(bc)}{(c+d)(b+d) \chi_{\text{máx.}}^2 - \chi_{\text{mín.}}^2}}
\end{aligned}$$

Si  $\chi_{\text{máx.}}^2 = n = a + b + c + d = 0 + b + c + d = b + c + d$  y

$$\begin{aligned}
\chi_{\text{mín.}}^2 = 0 \rightarrow A_{37}(j, k)^{(a=0,b,c,d)} &= \pm \sqrt{\frac{(b+c+d)(bc)}{(c+d)(b+d)} - 0} = \pm \sqrt{\frac{(b+c+d)(bc)}{(c+d)(b+d)}} \\
&= \pm \sqrt{\frac{(b+c+d)(bc)}{(c+d)(b+d)} \frac{1}{b+c+d}} = \pm \sqrt{\frac{bc}{(c+d)(b+d)}} \text{ NO ES EL MÍNIMO}
\end{aligned}$$

**Jaccard** [0, 1]

$$A_{38}(j, k)^{(a,b,c,d)} = \frac{a}{a + b + c}$$

$$A_{38}(j, k)^{(a=0,b,c,d)} = \frac{0}{0 + b + c} = \frac{0}{b + c} = 0 \text{ (mínimo)}$$

**Johnson\_1** [0, 2]

$$A_{39}(j, k)^{(a,b,c,d)} = \frac{a}{a + b} + \frac{a}{a + c}$$

$$A_{39}(j, k)^{(a=0,b,c,d)} = \frac{0}{0 + b} + \frac{0}{0 + c} = \frac{0}{b} + \frac{0}{c} = 0 \text{ (mínimo)}$$

**Johnson\_2** [0, +∞]

$$A_{40}(j, k)^{(a,b,c,d)} = \frac{a}{2b}$$

$$A_{40}(j, k)^{(a=0,b,c,d)} = \frac{0}{2b} = 0 \text{ (mínimo)}$$

**Kendall** [-1, 1]

$$A_{41}(j, k)^{(a,b,c,d)}$$

$$= \frac{2(ad - bc)}{\sqrt{(n^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)(n^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}} \\ = \frac{2(ad - bc)}{\sqrt{([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ab + 2cd)([a + b + c + d]^2 - a^2 - b^2 - c^2 - d^2 + 2ac + 2bd)}}$$

$$A_{41}(j, k)^{(a=0,b,c,d)}$$

$$= \frac{2(0d - bc)}{\sqrt{([0 + b + c + d]^2 - 0^2 - b^2 - c^2 - d^2 + 2b0 + 2cd)([0 + b + c + d]^2 - 0^2 - b^2 - c^2 - d^2 + 2c0 + 2bd)}} \\ = \frac{2(0d - bc)}{\sqrt{([b + c + d]^2 - b^2 - c^2 - d^2 + 2cd)([b + c + d]^2 - b^2 - c^2 - d^2 + 2bd)}} \\ = \frac{2(-bc)}{\sqrt{([b + c + d]^2 - b^2 - c^2 - d^2 + 2cd)([b + c + d]^2 - b^2 - c^2 - d^2 + 2bd)}} \\ = \frac{-2bc}{\sqrt{(\text{expresión}_1)(\text{expresión}_2)}}$$

$$\begin{aligned} \text{expresión}_1 &= [(b + c) + d]^2 - b^2 - c^2 - d^2 + 2cd \\ &= (b + c)^2 + 2(b + c)d + d^2 - b^2 - c^2 - d^2 + 2cd \\ &= b^2 + 2bc + c^2 + 2bd + 2cd + d^2 - b^2 - c^2 - d^2 + 2cd \\ &= 2bc + 2bd + 4cd \end{aligned}$$

$$\begin{aligned} \text{expresión}_2 &= [(b + c) + d]^2 - b^2 - c^2 - d^2 + 2bd \\ &= (b + c)^2 + 2(b + c)d + d^2 - b^2 - c^2 - d^2 + 2bd \\ &= b^2 + 2bc + c^2 + 2bd + 2cd + d^2 - b^2 - c^2 - d^2 + 2bd \\ &= 2bc + 2cd + 4bd \end{aligned}$$

$$A_{41}(j, k)^{(a=0,b,c,d)} = \frac{-2bd}{\sqrt{(2bc + 2bd + 4cd)(2bc + 2cd + 4bd)}} \quad \text{NO ES EL MÍNIMO}$$

**Kuder & Richardson**  $[-\infty, 1]$

$$A_{42}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a + b)(c + d) + (a + c)(b + d) + 2(ad - bc)}$$

$$\begin{aligned} A_{42}(j, k)^{(a=0,b,c,d)} &= \frac{4(0d - bc)}{(0 + b)(c + d) + (0 + c)(b + d) + 2(0d - bc)} \\ &= \frac{-4bc}{b(c + d) + c(b + d) + 2(-bc)} = \frac{-4bc}{bc + bd + cb + cd - 2bc} \\ &= \frac{-4bc}{2bc + bd + cd - 2bc} = \frac{-4bc}{bd + cd} = \frac{-4bc}{d(b + c)} \quad \text{NO ES EL MÍNIMO} \end{aligned}$$

**Kulczynski\_1**  $[0, +\infty]$

$$A_{43}(j, k)^{(a,b,c,d)} = \frac{a}{b + c}$$

$$A_{43}(j, k)^{(a=0,b,c,d)} = \frac{0}{b + c} = 0 \quad (\text{mínimo})$$

**Kulczynski\_2**  $[0, 1]$

$$A_{44}(j, k)^{(a,b,c,d)} = \frac{1}{2} \left( \frac{a}{a + b} + \frac{a}{a + c} \right)$$

$$A_{44}(j, k)^{(a=0,b,c,d)} = \frac{1}{2} \left( \frac{0}{0 + b} + \frac{0}{0 + c} \right) = \frac{1}{2} \left( \frac{0}{b} + \frac{0}{c} \right) = \frac{1}{2} * 0 = 0 \quad (\text{mínimo})$$

**Lamont & Grant**  $[0, 0.5]$

$$A_{45}(j, k)^{(a,b,c,d)} = \frac{a}{2a + b + c}$$

$$A_{45}(j, k)^{(a=0,b,c,d)} = \frac{0}{2 * 0 + b + c} = \frac{0}{b + c} = 0 \quad (\text{mínimo})$$

**Legendre & Legendre**  $[0, 1]$

$$A_{46}(j, k)^{(a,b,c,d)} = \frac{3a}{3a + b + c}$$

$$A_{46}(j, k)^{(a=0,b,c,d)} = \frac{3 * 0}{3 * 0 + b + c} = \frac{0}{b + c} = 0 \quad (\text{mínimo})$$

**Loevinger**  $[-\infty, 1]$

$$A_{47}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\text{mín.}\{(a+b)(b+d), (a+c)(c+d)\}}$$

$$A_{47}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{\text{mín.}\{(0+b)(b+d), (0+c)(c+d)\}} = \frac{-bc}{\text{mín.}\{b(b+d), c(c+d)\}}$$

$$\text{Si } b(b+d) < c(c+d) \rightarrow A_{47}(j, k)^{(a=0,b,c,d)} = \frac{-bc}{b(b+d)} = \frac{-c}{b+d} \text{ NO ES EL MÍNIMO}$$

$$\text{Si } b(b+d) \geq c(c+d) \rightarrow A_{47}(j, k)^{(a=0,b,c,d)} = \frac{-bc}{c(c+d)} = \frac{-b}{c+d} \text{ NO ES EL MÍNIMO}$$

### Maxwell & Pilliner [-1, 1]

$$A_{48}(j, k)^{(a,b,c,d)} = \frac{2(ad - bc)}{(a+b)(c+d) + (a+c)(b+d)}$$

$$A_{48}(j, k)^{(a=0,b,c,d)} = \frac{2(0d - bc)}{(0+b)(c+d) + (0+c)(b+d)} \\ = \frac{-2bc}{b(c+d) + c(b+d)} \text{ NO ES EL MÍNIMO}$$

### McConnaughey [-1, 1]

$$A_{49}(j, k)^{(a,b,c,d)} = \frac{a^2 - bc}{(a+b)(a+c)}$$

$$A_{49}(j, k)^{(a=0,b,c,d)} = \frac{0^2 - bc}{(0+b)(0+c)} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

### Michael [-1, 1]

$$A_{50}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc)}{(a+d)^2 + (b+c)^2}$$

$$A_{50}(j, k)^{(a=0,b,c,d)} = \frac{4(0d - bc)}{(0+d)^2 + (b+c)^2} = \frac{-4bc}{d^2 + (b+c)^2} \text{ NO ES EL MÍNIMO}$$

### Mountford [0, 2]

$$A_{51}(j, k)^{(a,b,c,d)} = \frac{a}{\frac{1}{2}(ab + ac) + bc}$$

$$A_{51}(j, k)^{(a=0,b,c,d)} = \frac{0}{\frac{1}{2}(0b + 0c) + bc} = \frac{0}{\frac{1}{2} * 0 + bc} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

### Ochiai [0, 1]

$$A_{52}(j, k)^{(a,b,c,d)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

$$A_{52}(j, k)^{(a=0,b,c,d)} = \frac{0}{\sqrt{(0+b)(0+c)}} = \frac{0}{\sqrt{bc}} = 0 \text{ (mínimo)}$$

**Pearson & Heron** [-1, 1]

$$A_{53}(j, k)^{(a,b,c,d)} = \cos \frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{53}(j, k)^{(a=0,b,c,d)} = \cos \frac{180\sqrt{bc}}{\sqrt{0d} + \sqrt{bc}} = \cos \frac{180\sqrt{bc}}{\sqrt{bc}} = \cos(180) = -1 \text{ (mínimo)}$$

**Pearson\_1** [0, n]

$$A_{54}(j, k)^{(a,b,c,d)} = \chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\begin{aligned} A_{54}(j, k)^{(a=0,b,c,d)} &= \frac{(0+b+c+d)(0d - bc)^2}{(0+b)(c+d)(0+c)(b+d)} = \frac{(b+c+d)(-bc)^2}{b(c+d)c(b+d)} = \frac{(b+c+d)(bc)^2}{bc(c+d)(b+d)} \\ &= \frac{bc(b+c+d)}{(c+d)(b+d)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Pearson\_2** [0, 1]

$$A_{55}(j, k)^{(a,b,c,d)} = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{\frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{(a+b+c+d) + \frac{(a+b+c+d)(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

$$\begin{aligned}
A_{55}(j, k)^{(a=0, b, c, d)} &= \sqrt{\frac{\frac{(0+b+c+d)(0d-bc)^2}{(0+b)(c+d)(0+c)(b+d)}}{(0+b+c+d) + \frac{(0+b+c+d)(0d-bc)^2}{(0+b)(c+d)(0+c)(b+d)}}} \\
&= \sqrt{\frac{\frac{(b+c+d)(bc)^2}{b(c+d)c(b+d)}}{(b+c+d) + \frac{(b+c+d)(bc)^2}{b(c+d)c(b+d)}}} = \sqrt{\frac{\frac{(b+c+d)(bc)^2}{bc(c+d)(b+d)}}{(b+c+d) + \frac{(b+c+d)(bc)^2}{bc(c+d)(b+d)}}} \\
&= \sqrt{\frac{\frac{(b+c+d)bc}{(c+d)(b+d)}}{(b+c+d) + \frac{(b+c+d)bc}{(c+d)(b+d)}}} \\
&= \sqrt{\frac{\frac{bc(b+c+d)}{(c+d)(b+d)}}{(b+c+d)(c+d)(b+d) + bc(b+c+d)}} \\
&= \sqrt{\frac{bc(b+c+d)(c+d)(b+d)}{(c+d)(b+d)(b+c+d)(c+d)(b+d) + bc(b+c+d)}} \\
&= \sqrt{\frac{bc(b+c+d)}{(b+c+d)(c+d)(b+d) + (b+c+d)bc}} \\
&= \sqrt{\frac{(b+c+d)bc}{(b+c+d)[(c+d)(b+d) + bc]}} \\
&= \sqrt{\frac{bc}{(c+d)(b+d) + bc}} \text{ NO ES EL MÍN.}
\end{aligned}$$

**Peirce\_1** [0, 1]

$$A_{56}(j, k)^{(a, b, c, d)} = \frac{ab + bc}{ab + 2bc + cd}$$

$$A_{56}(j, k)^{(a=0, b, c, d)} = \frac{0b + bc}{0b + 2bc + cd} = \frac{bc}{2bc + cd} = \frac{bc}{c(2b + d)} = \frac{b}{2b + d} \text{ NO ES EL MÍNIMO}$$

**Peirce\_2** [-1, 1]

$$A_{57}(j, k)^{(a, b, c, d)} = \frac{ad - bc}{(a+c)(b+d)}$$

$$A_{57}(j, k)^{(a=0, b, c, d)} = \frac{0d - bc}{(0+c)(b+d)} = \frac{-bc}{c(b+d)} = \frac{-b}{b+d} \text{ NO ES EL MÍNIMO}$$

**Peirce\_3** [-1, 1]

$$A_{58}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{(a + b)(c + d)}$$

$$A_{58}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{(0 + b)(c + d)} = \frac{-bc}{b(c + d)} = \frac{-c}{c + d} > -1 \text{ NO ES EL MÍNIMO}$$

**Roger & Tanimoto** [0, 1]

$$A_{59}(j, k)^{(a,b,c,d)} = \frac{a + d}{a + 2(b + c) + d}$$

$$A_{59}(j, k)^{(a=0,b,c,d)} = \frac{0 + d}{0 + 2(b + c) + d} = \frac{d}{2(b + c) + d} \text{ NO ES EL MÍNIMO}$$

**Rogot & Goldberg** [0, 1]

$$A_{60}(j, k)^{(a,b,c,d)} = \frac{a}{2a + b + c} + \frac{d}{2d + b + c}$$

$$\begin{aligned} A_{60}(j, k)^{(a=0,b,c,d)} &= \frac{0}{2 * 0 + b + c} + \frac{d}{2d + b + c} = \frac{0}{b + c} + \frac{d}{2d + b + c} \\ &= \frac{d}{2d + b + c} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Russell & Rao** [0, 1]

$$A_{61}(j, k)^{(a,b,c,d)} = \frac{a}{n} = \frac{a}{a + b + c + d}$$

$$A_{61}(j, k)^{(a=0,b,c,d)} = \frac{0}{0 + b + c + d} = \frac{0}{b + c + d} = 0 \text{ (mínimo)}$$

**Scott** [-1, 1]

$$A_{62}(j, k)^{(a,b,c,d)} = \frac{4(ad - bc) - (b - c)^2}{(2a + b + c)(2d + b + c)}$$

$$\begin{aligned} A_{62}(j, k)^{(a=0,b,c,d)} &= \frac{4(0d - bc) - (b - c)^2}{(2 * 0 + b + c)(2d + b + c)} = \frac{-4bc - (b - c)^2}{(b + c)(2d + b + c)} \\ &= \frac{-4bc - (b^2 - 2bc + c^2)}{2bd + b^2 + bc + 2cd + bc + c^2} = \frac{-b^2 - c^2 - 2bc}{b^2 + c^2 + 2bc + 2bd + 2cd} \\ &= \frac{-(b^2 + c^2 + 2bc)}{(b^2 + c^2 + 2bc) + 2bd + 2cd} = \frac{-(b + c)^2}{(b + c)^2 + 2d(b + c)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Simpson** [0, 1]

$$A_{63}(j, k)^{(a,b,c,d)} = \frac{a}{\text{mín.}\{(a + b), (a + c)\}}$$

$$A_{63}(j, k)^{(a=0, b, c, d)} = \frac{0}{\text{mín.}\{(0 + b), (0 + c)\}} = \frac{0}{\text{mín.}\{(b), (c)\}} = 0 \text{ (mínimo)}$$

**Sokal & Michener** [0, 1]

$$A_{64}(j, k)^{(a, b, c, d)} = \frac{a + d}{n} = \frac{a + d}{a + b + c + d}$$

$$A_{64}(j, k)^{(a=0, b, c, d)} = \frac{0 + d}{0 + b + c + d} = \frac{d}{b + c + d} = \frac{d}{d + (b + c)} \text{ NO ES EL MÍNIMO}$$

**Sokal & Sneath\_1** [0, 1]

$$A_{65}(j, k)^{(a, b, c, d)} = \frac{a}{a + 2(b + c)}$$

$$A_{65}(j, k)^{(a=0, b, c, d)} = \frac{0}{0 + 2(b + c)} = \frac{0}{2(b + c)} = 0 \text{ (mínimo)}$$

**Sokal & Sneath\_2** [0, 1]

$$A_{66}(j, k)^{(a, b, c, d)} = \frac{1}{4} \left( \frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{c + d} + \frac{d}{b + d} \right)$$

$$\begin{aligned} A_{66}(j, k)^{(a=0, b, c, d)} &= \frac{1}{4} \left( \frac{0}{0 + b} + \frac{0}{0 + c} + \frac{d}{c + d} + \frac{d}{b + d} \right) = \frac{1}{4} \left( \frac{0}{b} + \frac{0}{c} + \frac{d}{c + d} + \frac{d}{b + d} \right) \\ &= \frac{1}{4} \left( 0 + 0 + \frac{d}{c + d} + \frac{d}{b + d} \right) = \frac{1}{4} \left( \frac{d}{c + d} + \frac{d}{b + d} \right) \\ &= \frac{1}{4} \left( \frac{d(b + d) + d(c + d)}{(c + d)(b + d)} \right) = \frac{1}{4} \frac{d(b + d + c + d)}{(c + d)(b + d)} \\ &= \frac{d}{4} \frac{(b + c + 2d)}{(c + d)(b + d)} \text{ NO ES EL MÍNIMO} \end{aligned}$$

**Sokal & Sneath\_3** [0, +∞]

$$A_{67}(j, k)^{(a, b, c, d)} = \frac{a + d}{b + c}$$

$$A_{67}(j, k)^{(a=0, b, c, d)} = \frac{0 + d}{b + c} = \frac{d}{b + c} \text{ NO ES EL MÍNIMO}$$

**Sokal & Sneath\_4** [0, 1]

$$A_{68}(j, k)^{(a, b, c, d)} = \frac{a + d}{a + \frac{1}{2}(b + c) + d}$$

$$A_{68}(j, k)^{(a=0, b, c, d)} = \frac{0 + d}{0 + \frac{1}{2}(b + c) + d} = \frac{d}{\frac{1}{2}(b + c) + d} \text{ NO ES EL MÍNIMO}$$

**Sokal & Sneath\_5** [0, 1]

$$A_{69}(j, k)^{(a,b,c,d)} = \frac{ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$A_{69}(j, k)^{(a=0,b,c,d)} = \frac{0d}{\sqrt{(0+b)(c+d)(0+c)(b+d)}} = \frac{0}{\sqrt{b(c+d)c(b+d)}} = 0 \text{ (mínimo)}$$

**Sorensen** [0, 1]

$$A_{70}(j, k)^{(a,b,c,d)} = \frac{2a}{2a+b+c}$$

$$A_{70}(j, k)^{(a=0,b,c,d)} = \frac{2 * 0}{2 * 0 + b + c} = \frac{0}{b+c} = 0 \text{ (mínimo)}$$

**Sorgenfrei** [0, 1]

$$A_{71}(j, k)^{(a,b,c,d)} = \frac{a^2}{(a+b)(a+c)}$$

$$A_{71}(j, k)^{(a=0,b,c,d)} = \frac{0^2}{(0+b)(0+c)} = \frac{0}{bc} = 0 \text{ (mínimo)}$$

**Stiles**  $[-\infty, +\infty]$ 

$$A_{72}(j, k)^{(a,b,c,d)} = \log \frac{n(|ad - bc| - \frac{n}{2})^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \log \frac{(a+b+c+d)(|ad - bc| - \frac{a+b+c+d}{2})^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$A_{72}(j, k)^{(a=0,b,c,d)} = \log \frac{(0+b+c+d)(|0d - bc| - \frac{0+b+c+d}{2})^2}{(0+b)(0+c)(b+d)(c+d)}$$

$$= \log \frac{(b+c+d)(bc - \frac{b+c+d}{2})^2}{bc(b+d)(c+d)} = \log \frac{(b+c+d)(\frac{2bc - (b+c+d)}{2})^2}{bc(b+d)(c+d)}$$

$$= \log \frac{(b+c+d)(2bc - (b+c+d))^2}{2^2} \frac{1}{bc(b+d)(c+d)}$$

$$= \log \frac{(b+c+d)(2bc - (b+c+d))^2}{4bc(b+d)(c+d)} \text{ NO ES EL MÍNIMO}$$

**Tarwid** [-1, 1]

$$A_{73}(j, k)^{(a,b,c,d)} = \frac{na - (a+b)(a+c)}{na + (a+b)(a+c)} = \frac{(a+b+c+d)a - (a+b)(a+c)}{(a+b+c+d)a + (a+b)(a+c)}$$

$$A_{73}(j, k)^{(a=0,b,c,d)} = \frac{(0 + b + c + d)0 - (0 + b)(0 + c)}{(0 + b + c + d)0 + (0 + b)(0 + c)} = \frac{(b + c + d)0 - bc}{(b + c + d)0 + bc} = \frac{0 - bc}{0 + bc}$$

$$= \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Tschuproff** [0, 1]

$$A_{74}(j, k)^{(a,b,c,d)} = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}} = \sqrt{\frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}}$$

$$A_{74}(j, k)^{(a=0,b,c,d)} = \sqrt{\frac{(0 + b + c + d)(0d - bc)^2}{(0 + b)(c + d)(0 + c)(b + d)}}$$

$$= \sqrt{\frac{(b + c + d)(-bc)^2}{b(c + d)c(b + d)}} = \sqrt{\frac{(b + c + d)(bc)^2}{bc(c + d)(b + d)}} = \sqrt{\frac{(b + c + d)bc}{(c + d)(b + d)}}$$

$$= \sqrt{\frac{(b + c + d)bc}{(c + d)(b + d)} \cdot \frac{1}{(b + c + d)}} = \sqrt{\frac{bc}{(c + d)(b + d)}} \text{ NO ES EL MÍNIMO}$$

**Yule \_1** [-1, 1]

$$A_{75}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} = \phi$$

$$A_{75}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{\sqrt{(0 + b)(c + d)(0 + c)(b + d)}} = \frac{-bc}{\sqrt{b(c + d)c(b + d)}}$$

$$= \frac{-bc}{\sqrt{bc(c + d)(b + d)}} = \frac{-\sqrt{bc}\sqrt{bc}}{\sqrt{bc}\sqrt{(c + d)(b + d)}}$$

$$= \frac{-\sqrt{bc}}{\sqrt{(c + d)(b + d)}} \text{ NO ES EL MÍNIMO}$$

**Yule \_2** [-1, 1]

$$A_{76}(j, k)^{(a,b,c,d)} = \frac{ad - bc}{ad + bc}$$

$$A_{76}(j, k)^{(a=0,b,c,d)} = \frac{0d - bc}{0d + bc} = \frac{-bc}{bc} = -1 \text{ (mínimo)}$$

**Yule \_3** [-1, 1]

$$A_{77}(j, k)^{(a,b,c,d)} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

$$A_{77}(j, k)^{(a=0,b,c,d)} = \frac{\sqrt{0d} - \sqrt{bc}}{\sqrt{0d} + \sqrt{bc}} = \frac{-\sqrt{bc}}{\sqrt{bc}} = -1 \text{ (mínimo)}$$



## ANEXO 5

### Resultados propiedad de insesgamiento (12)

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	-0,936427	-0,850195	-0,696378	-0,007547	0,497499
		desvio	0,061124	0,098386	0,118119	0,191767	0,106481
	n=20	media	-0,935816	-0,849413	-0,696967	-0,015499	0,494824
		desvio	0,062186	0,100618	0,120397	0,195237	0,109198
<b>Anónimo_1</b>	n=50	media	-0,935279	-0,848367	-0,696548	-0,013102	0,494391
		desvio	0,062872	0,100643	0,120929	0,196214	0,109028
	n=100	media	-0,935230	-0,848527	-0,695965	-0,012986	0,495129
		desvio	0,063100	0,101379	0,120673	0,197940	0,109756
<b>media poblacional</b>			<b>-0,934426</b>	<b>-0,846154</b>	<b>-0,692308</b>	<b>0,000000</b>	<b>0,500000</b>
	n=10	media	0,574353	0,674263	0,590284	0,874503	0,831876
		desvio	0,048753	0,043329	0,049393	0,026650	0,035991
	n=20	media	0,573923	0,673881	0,590970	0,874201	0,831886
		desvio	0,049455	0,043642	0,049454	0,027305	0,036557
<b>Anónimo_2</b>	n=50	media	0,573773	0,674282	0,590967	0,874240	0,832226
		desvio	0,049839	0,044233	0,050132	0,027745	0,037040
	n=100	media	0,573958	0,674504	0,590965	0,874306	0,832152
		desvio	0,050036	0,044546	0,050227	0,028005	0,037082
<b>media poblacional</b>			<b>0,575540</b>	<b>0,675676</b>	<b>0,592593</b>	<b>0,875000</b>	<b>0,833333</b>
	n=10	media	0,148124	0,347695	0,182740	0,748657	0,663950
		desvio	0,097535	0,085867	0,097071	0,054848	0,071253
	n=20	media	0,147060	0,348427	0,180977	0,748782	0,664506
		desvio	0,098557	0,088177	0,099002	0,055025	0,072820
<b>Anónimo_3</b>	n=50	media	0,147360	0,347851	0,181842	0,748318	0,664360
		desvio	0,099319	0,088735	0,100684	0,055312	0,073705
	n=100	media	0,147550	0,348348	0,181852	0,748618	0,663625
		desvio	0,100187	0,088581	0,100682	0,055716	0,074372
<b>media poblacional</b>			<b>0,151079</b>	<b>0,351351</b>	<b>0,185185</b>	<b>0,750000</b>	<b>0,666667</b>
	n=10	media	0,058329	0,130422	0,235369	0,625487	0,787518
		desvio	0,053439	0,079261	0,075220	0,084671	0,042134
	n=20	media	0,058068	0,130683	0,235657	0,626437	0,786861
		desvio	0,054184	0,079905	0,076917	0,086600	0,043119
<b>Anónimo_4</b>	n=50	media	0,058614	0,130573	0,235391	0,625614	0,787175
		desvio	0,054594	0,080373	0,077962	0,087638	0,043056
	n=100	media	0,057901	0,130742	0,235581	0,625663	0,787006
		desvio	0,054344	0,080789	0,077935	0,087544	0,043303
<b>media poblacional</b>			<b>0,062040</b>	<b>0,138122</b>	<b>0,244275</b>	<b>0,636364</b>	<b>0,789474</b>
	n=10	media	-0,8926070	-0,1475900	-0,5029850	2,0334571	0,8894968
		desvio	0,1009995	0,5035446	0,1859903	0,5554379	0,1951394
	n=20	media	-0,8939570	-0,1565260	-0,5027520	2,0408947	0,8940092
		desvio	0,1028478	0,5185160	0,1840961	0,5674383	0,2012205
<b>Anónimo_5</b>	n=50	media	-0,8936590	-0,1487030	-0,5031340	2,0335535	0,8924481
		desvio	0,1035050	0,5215368	0,1875023	0,5702205	0,2010197
	n=100	media	-0,8931200	-0,1516110	-0,5027550	2,0368895	0,8920722
		desvio	0,1039214	0,5233601	0,1875786	0,5730398	0,2015030
<b>media poblacional</b>			<b>-0,8924730</b>	<b>-0,1489360</b>	<b>-0,5000000</b>	<b>1,9761905</b>	<b>0,8750000</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
<b>Baroni-Urbani &amp; Buser_1</b>	n=10	media	0,0863407	0,1817500	0,2520410	0,6403630	0,7736620
		desvio	0,0712725	0,0853230	0,0620260	0,0698990	0,0452120
	n=20	media	0,0864688	0,1803670	0,2519940	0,6402220	0,7730830
		desvio	0,0713195	0,0869400	0,0632070	0,0714720	0,0452840
	n=50	media	<b>0,0865860</b>	0,1801850	0,2515680	0,6400010	0,7731110
		desvio	0,0717984	0,0879380	0,0642550	0,0717210	0,0459650
	n=100	media	<b>0,0862373</b>	<b>0,1802900</b>	0,2514730	0,6403650	0,7729270
		desvio	0,0716617	0,0881870	0,0642560	0,0719000	0,0461160
	<b>media poblacional</b>		<b>0,1104350</b>	<b>0,2000000</b>	<b>0,2581816</b>	<b>0,6457513</b>	<b>0,7745967</b>
	<b>Baroni-Urbani &amp; Buser_2</b>	n=10	media	-0,8244590	-0,6379900	-0,4964590	0,2821195
desvio			0,1420800	0,1708845	0,1239965	0,1419008	0,0902222
n=20		media	-0,8267020	-0,6402690	-0,4959130	0,2820989	0,5468580
		desvio	0,1425110	0,1756333	0,1259215	0,1423841	0,0901724
n=50		media	<b>-0,8265060</b>	-0,6417980	-0,4963230	0,2816788	0,5466615
		desvio	0,1429633	0,1766444	0,1280807	0,1429188	0,0917440
n=100		media	<b>-0,8274430</b>	<b>-0,6404880</b>	-0,4973500	0,2806317	0,5467610
		desvio	0,1433435	0,1769334	0,1285500	0,1441751	0,0918612
<b>media poblacional</b>		<b>-0,7791300</b>	<b>-0,6000000</b>	<b>-0,4836370</b>	<b>0,2915026</b>	<b>0,5491933</b>	
<b>Braun-Blanquet</b>		n=10	media	0,0294928	0,0427792	0,1251699	0,3565391
	desvio		0,0281427	0,0286988	0,0504168	0,0881516	0,0589214
	n=20	media	0,0296143	0,0425587	0,1249608	0,3565301	<b>0,7167466</b>
		desvio	0,0287118	0,0291479	0,0519840	0,0902372	0,0593100
	n=50	media	0,0294765	0,0425217	0,1251238	0,3571519	<b>0,7168831</b>
		desvio	0,0288900	0,0294946	0,0518670	0,0905851	0,0604355
	n=100	media	0,0295522	0,0426675	0,1248501	0,3568299	<b>0,7163910</b>
		desvio	0,0288812	0,0295109	0,0522676	0,0913791	0,0604616
	<b>media poblacional</b>		<b>0,0322581</b>	<b>0,0425532</b>	<b>0,1250000</b>	<b>0,3571429</b>	<b>0,7500000</b>
	<b>Clement</b>	n=10	media	19,4934850	40,3105790	28,3333850	82,8545300
desvio			3,7649032	23,2473880	6,3342198	9,7918115	4,5305361
n=20		media	19,4320140	40,8318070	28,3504660	82,8858270	78,3026270
		desvio	3,7977176	23,4783330	6,3621255	9,8761569	4,6463356
n=50		media	19,4598880	40,5688530	28,3133510	82,9864900	78,3259420
		desvio	3,8510160	23,5937830	6,4749102	9,8858990	4,6471951
n=100		media	19,4723390	40,5869770	28,3136890	82,8891650	78,3282180
		desvio	3,8555707	23,7025050	6,4644896	9,9624545	4,6226284
<b>media poblacional</b>		<b>19,4761900</b>	<b>40,6315790</b>	<b>28,3333330</b>	<b>82,8787880</b>	<b>78,3333330</b>	
<b>Clifford &amp; Stephenson</b>		n=10	media	0,1571230	<b>0,0106580</b>	0,0631259	0,2141755
	desvio		0,0461829	0,0120420	0,0401127	0,0864451	0,0940528
	n=20	media	0,1564768	<b>0,0108160</b>	0,0628331	0,2148339	0,3470057
		desvio	0,0475233	0,0125136	0,0408937	0,0869620	0,0953572
	n=50	media	0,1564310	<b>0,0107265</b>	0,0628449	0,2142289	0,3480067
		desvio	0,0477310	0,0127438	0,0412885	0,0875853	0,0960735
	n=100	media	0,1565060	<b>0,0108034</b>	0,0629842	0,2141578	0,3472851
		desvio	0,0479090	0,0128960	0,0414618	0,0877200	0,0967485
	<b>media poblacional</b>		<b>0,1533649</b>	<b>0,0010353</b>	<b>0,0555556</b>	<b>0,2071008</b>	<b>0,3402778</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	-0,3835570	-0,0152700	-0,2201650	0,3934477	0,5816661
		desvio	0,0603152	0,0447357	0,0810032	0,0994526	0,0805035
	n=20	media	-0,3846190	-0,0145390	-0,2186430	0,3956743	0,5816431
		desvio	0,0603791	0,0450178	0,0823532	0,1015057	0,0821107
<b>Cohen</b>	n=50	media	-0,3838920	-0,0148600	-0,2192640	0,3951001	0,5804389
		desvio	0,0617353	0,0455925	0,0834602	0,1021673	0,0827393
	n=100	media	-0,3842790	-0,0147060	-0,2190190	0,3952373	0,5804596
		desvio	0,0619924	0,0456083	0,0837184	0,1024037	0,0831051
<b>media poblacional</b>			<b>-0,3915090</b>	<b>-0,0147990</b>	<b>-0,2222220</b>	<b>0,3990385</b>	<b>0,5833333</b>
	n=10	media	-0,5111870	-0,0456320	-0,3188080	0,5761383	0,7083975
		desvio	0,0678524	0,1364006	0,1130766	0,1041527	0,0747920
	n=20	media	-0,5124440	-0,0428280	-0,3213430	0,5779050	0,7080274
		desvio	0,0696866	0,1379670	0,1123111	0,1053958	0,0759594
<b>Cole_1</b>	n=50	media	-0,5123440	-0,0428620	-0,3205280	0,5771414	0,7084132
		desvio	0,0701774	0,1376600	0,1140020	0,1068984	0,0761932
	n=100	media	-0,5123380	-0,0433000	-0,3198320	0,5775379	0,7086130
		desvio	0,0706433	0,1383024	0,1141914	0,1067575	0,0767034
<b>media poblacional</b>			<b>-0,5156970</b>	<b>-0,0454800</b>	<b>-0,3244430</b>	<b>0,5857794</b>	<b>0,7125812</b>
	n=10	media	0,8925171	<b>0,4124956</b>	0,5013981	0,7677693	0,6292886
		desvio	0,1000586	0,3098235	0,1774482	0,1452885	0,0911453
	n=20	media	0,8937933	<b>0,4153338</b>	0,5031913	0,7713698	0,6286951
		desvio	0,1021897	0,3163581	0,1832133	0,1482552	0,0926926
<b>Cole_2</b>	n=50	media	0,8937935	<b>0,4118856</b>	0,5025668	0,7698048	0,6281425
		desvio	0,1029712	0,3163574	0,1832631	0,1488124	0,0939040
	n=100	media	0,8933503	<b>0,4147451</b>	0,5033166	0,7701257	0,6292461
		desvio	0,1036945	0,3179390	0,1848993	0,1497251	0,0939650
<b>media poblacional</b>			<b>0,8924731</b>	<b>0,1489362</b>	<b>0,5000000</b>	<b>0,7685185</b>	<b>0,5833333</b>
	n=10	media	-0,4069230	-0,1349150	-0,3386900	0,7700623	0,5854223
		desvio	0,0953642	0,4610059	0,1332024	0,1451552	0,0969386
	n=20	media	-0,4056690	-0,1388420	-0,3383760	0,7721683	0,5848200
		desvio	0,0966695	0,4614520	0,1369563	0,1481230	0,0976018
<b>Cole_3</b>	n=50	media	-0,4075350	-0,1325400	-0,3376900	0,7698586	0,5860777
		desvio	0,0980470	0,4674782	0,1378152	0,1501848	0,0992151
	n=100	media	-0,4060010	-0,1330460	-0,3375960	0,7704220	0,5856647
		desvio	0,0975454	0,4639764	0,1389247	0,1501722	0,0998406
<b>media poblacional</b>			<b>-0,4009660</b>	<b>-0,1320750</b>	<b>-0,3333330</b>	<b>0,7685185</b>	<b>0,5833333</b>
	n=10	media	-0,3875040	-0,0081020	-0,1684320	0,2718133	0,5839286
		desvio	0,0909686	0,0239542	0,0688715	0,0846985	0,0974954
	n=20	media	-0,3874590	-0,0080570	-0,1684630	0,2714334	0,5857412
		desvio	0,0932091	0,0241815	0,0706971	0,0841065	0,0982762
<b>Cole_4</b>	n=50	media	-0,3877480	-0,0080480	-0,1686040	0,2711007	0,5857708
		desvio	0,0940081	0,0244738	0,0710087	0,0850698	0,1000168
	n=100	media	-0,3878610	-0,0079000	-0,1688890	0,2716784	0,5856810
		desvio	0,0941915	0,0245956	0,0712067	0,0853626	0,0996172
<b>media poblacional</b>			<b>-0,3824880</b>	<b>-0,0078390</b>	<b>-0,1666670</b>	<b>0,2694805</b>	<b>0,5833333</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	-2,6967940	-0,2256770	-1,5768350	3,6024857	3,4847092
		desvio	0,3585940	0,6877328	0,5621843	0,7597833	0,5347463
	n=20	media	-2,7010750	-0,2178160	-1,5729170	3,5955339	3,4906131
		desvio	0,3629654	0,7023654	0,5691972	0,7720890	0,5417214
<b>Dennis</b>	n=50	media	-2,6989580	-0,2228600	-1,5688660	3,5903943	3,4907933
		desvio	0,3661930	0,6966618	0,5781365	0,7777497	0,5469795
	n=100	media	-2,7006680	-0,2232420	-1,5706700	3,5896817	3,4921585
		desvio	0,3684284	0,6974649	0,5773008	0,7770700	0,5469107
<b>media poblacional</b>			<b>-2,7216770</b>	<b>-0,2283150</b>	<b>-1,5811390</b>	<b>3,6224170</b>	<b>3,5000000</b>
	n=10	media	0,0335484	0,3944288	0,1999856	0,8335600	0,7493225
		desvio	0,0319888	0,2396148	0,0782377	0,1082347	0,0669921
	n=20	media	0,0333400	0,3987995	0,1998610	0,8324631	0,7496319
		desvio	0,0325020	0,2446073	0,0797232	0,1112222	0,0684249
<b>Dice_1</b>	n=50	media	0,0330410	0,4007172	0,2005534	0,8341158	0,7500430
		desvio	0,0326010	0,2455668	0,0813965	0,1111780	0,0685402
	n=100	media	0,0333060	0,4005816	0,2001316	0,8331949	0,7499258
		desvio	0,0331600	0,2459154	0,0809100	0,1117089	0,0687298
<b>media poblacional</b>			<b>0,0333333</b>	<b>0,4000000</b>	<b>0,2000000</b>	<b>0,8333333</b>	<b>0,7500000</b>
	n=10	media	0,0321302	0,0420848	0,1253209	0,3572549	0,7491433
		desvio	0,0310843	0,0286282	0,0505634	0,0896557	0,0674695
	n=20	media	0,0323649	0,0423325	0,1251369	0,3572890	0,7496438
		desvio	0,0316135	0,0292892	0,0517214	0,0906954	0,0683294
<b>Dice_2</b>	n=50	media	0,0322789	0,0426696	0,1250415	0,3577095	0,7498770
		desvio	0,0319066	0,0295008	0,0526603	0,0910254	0,0688035
	n=100	media	0,0322592	0,0425184	0,1247737	0,3571257	0,7500763
		desvio	0,0319896	0,0295574	0,0525365	0,0918794	0,0687969
<b>media poblacional</b>			<b>0,0322581</b>	<b>0,0425532</b>	<b>0,1250000</b>	<b>0,3571429</b>	<b>0,7500000</b>
	n=10	media	0,0021192	0,0169277	0,0099660	0,0607008	0,0378760
		desvio	0,0020117	0,0101704	0,0036851	0,0107950	0,0040228
	n=20	media	0,0021334	0,0169977	0,0099857	0,0607243	0,0378818
		desvio	0,0020687	0,0103148	0,0036903	0,0111615	0,0040357
<b>Dice_3</b>	n=50	media	0,0021453	0,0170209	0,0099427	0,0608789	0,0378785
		desvio	0,0020751	0,0104578	0,0037325	0,0114261	0,0040499
	n=100	media	0,0021332	0,0170040	0,0099280	0,0607476	0,0378763
		desvio	0,0020779	0,0104325	0,0037270	0,0115135	0,0040856
<b>media poblacional</b>			<b>0,0021505</b>	<b>0,0170213</b>	<b>0,0100000</b>	<b>0,0595238</b>	<b>0,0375000</b>
	n=10	media	-0,8430680	-0,1404230	-0,4473790	0,8111468	0,7675005
		desvio	0,1362833	0,4569614	0,1695323	0,1177690	0,0757059
	n=20	media	-0,8446320	-0,1416950	-0,4452700	0,8090675	0,7683151
		desvio	0,1376217	0,4571137	0,1718360	0,1207648	0,0775548
<b>Digby</b>	n=50	media	-0,8448710	-0,1481760	-0,4462520	0,8101066	0,7681666
		desvio	0,1390844	0,4623192	0,1728267	0,1208461	0,0789159
	n=100	media	-0,8447960	-0,1445060	-0,4470940	0,8108446	0,7683645
		desvio	0,1388830	0,4663583	0,1731612	0,1222971	0,0789954
<b>media poblacional</b>			<b>-0,8193560</b>	<b>-0,1120670</b>	<b>-0,4380260</b>	<b>0,8050624</b>	<b>0,7680350</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30	
		b	0,29	0,03	0,20	0,02	0,10	
		c	0,30	0,45	0,35	0,18	0,10	
		d	0,40	0,50	0,40	0,70	0,50	
<b>Doolittle</b>	n=10	media	0,1555821	<b>0,0108805</b>	0,0626634	0,2149103	0,3487489	
		desvio	0,0464658	0,0121143	0,0399879	0,0855739	0,0958518	
	n=20	media	0,1562540	<b>0,0107097</b>	0,0627084	0,2138958	0,3458996	
		desvio	0,0478290	0,0123219	0,0406713	0,0875883	0,0949282	
	n=50	media	0,1565876	<b>0,0108061</b>	0,0630435	0,2140915	0,3468872	
		desvio	0,0477692	0,0127697	0,0410217	0,0874452	0,0963499	
	n=100	media	0,1563312	<b>0,0107653</b>	0,0628683	0,2150160	0,3471734	
		desvio	0,0479326	0,0128312	0,0412161	0,0881574	0,0960300	
	<b>media poblacional</b>		<b>0,1533649</b>	<b>0,0010353</b>	<b>0,0555556</b>	<b>0,2071008</b>	<b>0,3402778</b>	
	<b>Eyraud</b>	n=10	media	-0,0002080	-0,0001990	-0,0002230	-0,0001540	-0,0002780
			desvio	0,0000150	0,0000191	0,0000193	0,0000135	0,0000408
		n=20	media	-0,0002080	-0,0001990	-0,0002230	-0,0001540	-0,0002770
desvio			0,0000152	0,0000194	0,0000195	0,0000140	0,0000419	
n=50		media	-0,0002080	-0,0001990	-0,0002230	-0,0001540	-0,0002770	
		desvio	0,0000154	0,0000195	0,0000198	0,0000139	0,0000419	
n=100		media	-0,0002080	-0,0001990	-0,0002230	-0,0001540	-0,0002770	
		desvio	0,0000154	0,0000196	0,0000200	0,0000141	0,0000420	
<b>media poblacional</b>		<b>-0,0002070</b>	<b>-0,0001970</b>	<b>-0,0002210</b>	<b>-0,0001530</b>	<b>-0,0002730</b>		
<b>Fager &amp; McGowan</b>		n=10	media	-0,0542380	0,0548525	0,0781680	0,4440546	0,6716547
			desvio	0,0320978	0,0741183	0,0606173	0,0886195	0,0547406
		n=20	media	-0,0543610	0,0541784	0,0773930	0,4465995	0,6707960
	desvio		0,0321334	0,0757537	0,0621256	0,0895155	0,0555352	
	n=50	media	-0,0538430	0,0544713	0,0778558	0,4458933	0,6707638	
		desvio	0,0327924	0,0759439	0,0623872	0,0904439	0,0563638	
	n=100	media	-0,0541200	0,0541705	0,0775216	0,4454980	0,6712629	
		desvio	0,0327854	0,0763651	0,0628716	0,0906560	0,0562415	
	<b>media poblacional</b>		<b>-0,0570110</b>	<b>0,0575331</b>	<b>0,0790569</b>	<b>0,4510536</b>	<b>0,6709431</b>	
	<b>Faith</b>	n=10	media	0,2099820	0,2698635	0,2499540	0,4502410	0,5501245
			desvio	0,0247419	0,0262161	0,0289404	0,0263024	0,0340616
		n=20	media	0,2099038	0,2700170	0,2500823	0,4499827	0,5501930
desvio			0,0252900	0,0262926	0,0290338	0,0265532	0,0343744	
n=50		media	0,2098733	0,2700636	0,2499888	0,4500464	0,5502366	
		desvio	0,0255204	0,0267922	0,0292934	0,0268777	0,0348493	
n=100		media	0,2099150	0,2699369	0,2498526	0,4500452	0,5498476	
		desvio	0,0257084	0,0267643	0,0294800	0,0268892	0,0347668	
<b>media poblacional</b>		<b>0,2100000</b>	<b>0,2700000</b>	<b>0,2500000</b>	<b>0,4500000</b>	<b>0,5500000</b>		
<b>Fleiss</b>		n=10	media	-0,7960380	-0,1424450	-0,5051510	1,0393394	1,1733638
			desvio	0,1227823	0,4791246	0,1846764	0,1950007	0,1602232
		n=20	media	-0,7955050	-0,1363840	-0,5054380	1,0422729	1,1704495
	desvio		0,1243671	0,4842427	0,1877798	0,1977263	0,1639982	
	n=50	media	-0,7948110	-0,1398840	-0,5055100	1,0421816	1,1719372	
		desvio	0,1251997	0,4861944	0,1900128	0,2005570	0,1655025	
	n=100	media	-0,7947950	-0,1420380	-0,5068020	1,0415646	1,1718007	
		desvio	0,1252010	0,4874259	0,1906252	0,2017757	0,1664961	
	<b>media poblacional</b>		<b>-0,7834550</b>	<b>-0,1399140</b>	<b>-0,5000000</b>	<b>1,0379990</b>	<b>1,1666667</b>	

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	0,1062788	0,8472675	0,5002806	3,0234627	1,8923445
		desvio	0,1017396	0,5044684	0,1848898	0,5480911	0,1967035
	n=20	media	0,1074761	0,8516987	0,4957499	3,0323718	1,8911239
		desvio	0,1024950	0,5158976	0,1828856	0,5624460	0,1984596
<b>Forbes_1</b>	n=50	media	0,1067965	0,8507269	0,4972767	3,0405355	1,8917901
		desvio	0,1036632	0,5195602	0,1867780	0,5713294	0,2019248
	n=100	media	0,1064630	0,8510387	0,4976334	3,0399988	1,8916749
		desvio	0,1038407	0,5207762	0,1870010	0,5776890	0,2028971
<b>media poblacional</b>			<b>0,1075269</b>	<b>0,8510638</b>	<b>0,5000000</b>	<b>2,9761905</b>	<b>1,8750000</b>
	n=10	media	<b>-0,3381920</b>	-0,0083260	-0,1680020	0,2712191	0,5424420
		desvio	0,0638947	0,0237373	0,0680033	0,0828576	0,0831116
	n=20	media	<b>-0,3398660</b>	-0,0079930	-0,1684420	0,2723151	<b>0,5427423</b>
		desvio	0,0657656	0,0242113	0,0696200	0,0838939	0,0841154
<b>Forbes_2</b>	n=50	media	<b>-0,3395540</b>	-0,0080890	-0,1680250	0,2713422	<b>0,5429723</b>
		desvio	0,0657693	0,0244918	0,0704384	0,0849185	0,0852321
	n=100	media	<b>-0,3398360</b>	-0,0079560	-0,1682760	0,2714846	<b>0,5426344</b>
		desvio	0,0664112	0,0245925	0,0704466	0,0854360	0,0852095
<b>media poblacional</b>			<b>-0,3824880</b>	<b>-0,0078390</b>	<b>-0,1666670</b>	<b>0,2694805</b>	<b>0,5833333</b>
	n=10	media	-0,8929620	-0,1511450	-0,5025710	2,0396449	0,8911626
		desvio	0,1019025	0,5015785	0,1840926	0,5522861	0,1926390
	n=20	media	-0,8943770	-0,1523900	-0,5019570	2,0413764	0,8930998
		desvio	0,1021150	0,5176515	0,1848699	0,5668536	0,2009027
<b>Forbes_3</b>	n=50	media	-0,8934320	-0,1531840	-0,5024940	2,0396989	0,8932828
		desvio	0,1039838	0,5208317	0,1857952	0,5724684	0,2023799
	n=100	media	-0,8934250	-0,1499590	-0,5028040	2,0376607	0,8919374
		desvio	0,1040351	0,5199473	0,1872630	0,5758920	0,2025096
<b>media poblacional</b>			<b>-0,8924730</b>	<b>-0,1489360</b>	<b>-0,5000000</b>	<b>1,9761905</b>	<b>0,8750000</b>
	n=10	media	0,1293035	1,4758944	2,3647187	27,1967680	54,5710140
		desvio	0,1901867	1,5830901	1,7923217	9,1011187	7,8447980
	n=20	media	<b>0,1269306</b>	1,4874775	2,3571073	27,1213340	54,5414820
		desvio	0,2079198	1,6578614	1,8426815	9,1503554	7,8788324
<b>Fossum</b>	n=50	media	<b>0,1291539</b>	<b>1,4848894</b>	2,3816827	27,0783140	54,4700190
		desvio	0,2238493	1,6897934	1,9008438	9,2524215	7,9891498
	n=100	media	<b>0,1291443</b>	<b>1,4947693</b>	2,3843018	27,1566460	54,4794610
		desvio	0,2306718	1,7095552	1,9001047	9,2372147	8,0438996
<b>media poblacional</b>			<b>0,0268817</b>	<b>0,9574468</b>	<b>2,0250000</b>	<b>26,8601190</b>	<b>54,3906250</b>
	n=10	media	<b>-1,8865110</b>	-0,1279410	-0,7763030	1,0951492	0,6311328
		desvio	0,3761716	0,4595900	0,4146679	0,1812798	0,1034710
	n=20	media	<b>-1,8851750</b>	-0,1265220	-0,7715240	1,0930020	0,6328336
		desvio	0,4282034	0,4587959	0,4201597	0,1821782	0,1038130
<b>Gilbert &amp; Wells</b>	n=50	media	<b>-1,8836430</b>	-0,1299420	-0,7698710	1,0926925	0,6314342
		desvio	0,4327348	0,4657542	0,4217585	0,1836573	0,1047376
	n=100	media	<b>-1,8863090</b>	-0,1295070	-0,7710330	1,0943914	0,6318768
		desvio	0,4352967	0,4651777	0,4260827	0,1851810	0,1058132
<b>media poblacional</b>			<b>-2,2300140</b>	<b>-0,1612680</b>	<b>-0,6931470</b>	<b>1,0906441</b>	<b>0,6286087</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	0,4419585	0,5131513	0,4676414	0,7067331	0,7063728
		desvio	0,0308167	0,0312182	0,0313152	0,0308372	0,0310022
	n=20	media	0,4420121	0,5128902	0,4680926	0,7058746	0,7060565
		desvio	0,0314673	0,0316065	0,0315038	0,0318205	0,0316316
<b>Goodall</b>	n=50	media	0,4421101	0,5127935	0,4677872	0,7061018	0,7059976
		desvio	0,0317742	0,0318840	0,0317890	0,0321072	0,0319806
	n=100	media	0,4419403	0,5129039	0,4678827	0,7059153	0,7060518
		desvio	0,0319827	0,0319077	0,0318104	0,0321269	0,0320371
<b>media poblacional</b>			<b>0,4423902</b>	<b>0,5127358</b>	<b>0,4681157</b>	<b>0,7048328</b>	<b>0,7048328</b>
	n=10	media	<b>-0,8993610</b>	-0,8262670	-0,6952760	-0,0145250	0,4948449
		desvio	0,0447553	0,0849402	0,1164852	0,1944595	0,1071287
<b>Goodman &amp; Kruskal</b>	n=20	media	<b>-0,8978620</b>	-0,8256930	-0,6936970	-0,0129130	0,4934783
		desvio	0,0477845	0,0876145	0,1182135	0,1967708	0,1083769
	n=50	media	<b>-0,8980170</b>	-0,8256700	-0,6931640	-0,0147860	0,4957208
		desvio	0,0495181	0,0883006	0,1187383	0,1967338	0,1083063
	n=100	media	<b>-0,8980450</b>	<b>-0,8261630</b>	-0,6937850	-0,0139240	0,4952270
		desvio	0,0500032	0,0879963	0,1188570	0,1974505	0,1098638
<b>media poblacional</b>			<b>-0,9344260</b>	<b>-0,8461540</b>	<b>-0,6923080</b>	<b>0,0000000</b>	<b>0,5000000</b>
	n=10	media	-0,1789840	0,0396120	-0,1006520	0,5995040	0,5993560
		desvio	0,0952181	0,0969377	0,0975183	0,0784831	0,0774695
	n=20	media	-0,1801950	0,0395720	-0,0999320	0,6001030	0,6002530
		desvio	0,0961596	0,0986520	0,0979430	0,0790485	0,0794340
<b>Hamann</b>	n=50	media	-0,1804470	0,0404252	-0,1005230	0,5999168	0,6002848
		desvio	0,0977947	0,0993038	0,0990565	0,0796710	0,0791522
	n=100	media	-0,1805000	0,0398260	-0,0994770	0,5999528	0,5998244
		desvio	0,0984783	0,0996676	0,0989314	0,0797119	0,0796033
<b>media poblacional</b>			<b>-0,1800000</b>	<b>0,0400000</b>	<b>-0,1000000</b>	<b>0,6000000</b>	<b>0,6000000</b>
	n=10	media	13,3538730	16,0752740	19,0974400	42,1127500	64,4322770
		desvio	1,5115729	2,4825254	3,0814747	7,9039013	5,7143661
<b>Harris &amp; Lahey</b>	n=20	media	13,3514020	16,0787380	19,1806060	41,9767200	64,2883490
		desvio	1,5425335	2,5563292	3,1349183	7,8524755	5,8667502
	n=50	media	13,3623150	16,0850990	19,1326160	41,9964180	64,3523390
		desvio	1,5602508	2,5612590	3,1387622	7,9831111	5,8951569
	n=100	media	13,3518890	16,0645740	19,1591160	42,0071390	64,3518180
		desvio	1,5650449	2,5689433	3,1651369	8,0510551	5,9196270
<b>media poblacional</b>			<b>13,4815660</b>	<b>16,2253060</b>	<b>19,3092110</b>	<b>42,2222220</b>	<b>64,5714290</b>
	n=10	media	0,2104813	0,2752795	0,2519465	0,5556532	0,6573343
		desvio	0,0257115	0,0302332	0,0321962	0,0565460	0,0545743
<b>Hawkins &amp; Dotson</b>	n=20	media	0,2104281	0,2752610	0,2523728	0,5551161	0,6569851
		desvio	0,0264281	0,0300893	0,0330329	0,0570974	0,0555721
	n=50	media	0,2102025	0,2749709	0,2520181	0,5556907	0,6572647
		desvio	0,0264536	0,0303221	0,0332971	0,0577127	0,0560827
	n=100	media	0,2103355	0,2750641	0,2521754	0,5559451	0,6570304
		desvio	0,0266483	0,0304553	0,0337116	0,0577178	0,0561079
<b>media poblacional</b>			<b>0,2103535</b>	<b>0,2751020</b>	<b>0,2521930</b>	<b>0,5555556</b>	<b>0,6571429</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30	
		b	0,29	0,03	0,20	0,02	0,10	
		c	0,30	0,45	0,35	0,18	0,10	
		d	0,40	0,50	0,40	0,70	0,50	
<b>Hurlbert_1</b>	n=10	media	0,4539247	<b>0,3810564</b>	0,3382825	0,7676869	0,6292886	
		desvio	0,0832027	0,2881707	0,1309657	0,1453773	0,0911453	
	n=20	media	<b>0,4546016</b>	<b>0,3833365</b>	0,3387775	0,7713698	<b>0,6286951</b>	
		desvio	0,0865030	0,2953016	0,1356202	0,1482552	0,0926926	
	n=50	media	<b>0,4554667</b>	<b>0,3796659</b>	0,3387510	0,7697819	<b>0,6281425</b>	
		desvio	0,0865076	0,2943771	0,1352336	0,1488502	0,0939040	
	n=100	media	<b>0,4548974</b>	<b>0,3824402</b>	0,3392769	0,7700814	<b>0,6292461</b>	
		desvio	0,0873104	0,2965694	0,1363788	0,1497655	0,0939650	
	<b>media poblacional</b>		<b>0,4009662</b>	<b>0,1320755</b>	<b>0,3333333</b>	<b>0,7685185</b>	<b>0,5833333</b>	
	<b>Hurlbert_2</b>	n=10	media	-0,2138840	<b>-0,3953570</b>	-0,8036260	0,7020067	-0,3707640
			desvio	0,1782518	1,6506780	0,5249268	0,2009970	0,7680219
		n=20	media	-0,2102540	<b>-0,4237500</b>	-0,7906470	0,7059985	-0,3500860
desvio			0,1796112	1,6646707	0,5665472	0,2085551	0,7915591	
n=50		media	-0,2114150	<b>-0,4090560</b>	-0,7938320	0,7039779	-0,3612620	
		desvio	0,1816688	1,6924758	0,5960740	0,2131785	0,8035414	
n=100		media	-0,2113240	<b>-0,4075550</b>	-0,7976190	0,7038302	-0,3588690	
		desvio	0,1815145	1,6960393	0,6076788	0,2161822	0,8123763	
<b>media poblacional</b>		<b>-0,2268520</b>	<b>-1,8973670</b>	<b>-0,7745970</b>	<b>0,7194976</b>	<b>-0,4330130</b>		
<b>Jaccard</b>		n=10	media	0,0169527	0,0400302	0,0828339	0,3340781	0,6003316
			desvio	0,0162062	0,0270054	0,0346595	0,0851774	0,0681670
		n=20	media	0,0168524	0,0399940	0,0837612	0,3333355	0,6001679
	desvio		0,0163131	0,0273930	0,0354312	0,0850238	0,0690672	
	n=50	media	0,0166351	0,0401004	0,0831121	0,3339684	0,6003421	
		desvio	0,0163798	0,0277027	0,0356584	0,0873436	0,0696396	
	n=100	media	0,0166825	0,0400691	0,0835093	0,3333554	0,6000739	
		desvio	0,0165160	0,0278164	0,0356988	0,0871347	0,0696421	
	<b>media poblacional</b>		<b>0,0166667</b>	<b>0,0400000</b>	<b>0,0833333</b>	<b>0,3333333</b>	<b>0,6000000</b>	
	<b>Johnson_1</b>	n=10	media	0,0650363	0,4387291	0,3259889	1,1937252	1,5001401
			desvio	0,0613779	0,2641815	0,1255934	0,1598388	0,1059408
		n=20	media	0,0654716	0,4440051	0,3231569	1,1901753	1,4992901
desvio			0,0632142	0,2659236	0,1262393	0,1642572	0,1083149	
n=50		media	0,0657489	0,4431824	0,3267548	1,1917633	1,5003944	
		desvio	0,0642453	0,2664236	0,1281169	0,1649365	0,1087282	
n=100		media	0,0656991	0,4428707	0,3253367	1,1900792	1,4991237	
		desvio	0,0640073	0,2678269	0,1281767	0,1661871	0,1090979	
<b>media poblacional</b>		<b>0,0655914</b>	<b>0,4425532</b>	<b>0,3250000</b>	<b>1,1904762</b>	<b>1,5000000</b>		
<b>Johnson_2</b>		n=10	media	0,0177447	0,4316133	0,1313667	2,9174146	1,6785964
			desvio	0,0176227	0,4043003	0,0652527	1,7613641	0,6963355
		n=20	media	0,0179063	0,4312211	0,1314654	2,9025960	1,6924107
	desvio		0,0181277	0,4282550	0,0665472	1,7748100	0,7459757	
	n=50	media	0,0179161	0,4397516	0,1313246	2,8901011	1,6897584	
		desvio	0,0182439	0,4495116	0,0677599	1,7978228	0,7648548	
	n=100	media	0,0179476	<b>0,4336560</b>	0,1322001	<b>2,8919297</b>	<b>1,6913668</b>	
		desvio	0,0184555	0,4476486	0,0686637	1,8107975	0,7814709	
	<b>media poblacional</b>		<b>0,0172414</b>	<b>0,3333333</b>	<b>0,1250000</b>	<b>2,5000000</b>	<b>1,5000000</b>	

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	-0,1829530	-0,0089050	-0,1089740	0,2167035	0,3541150
		desvio	0,0354505	0,0273042	0,0414492	0,0655266	0,0698029
	n=20	media	-0,1830070	-0,0090640	-0,1094840	0,2163479	0,3544767
		desvio	0,0355820	0,0276443	0,0422411	0,0664295	0,0711904
<b>Kendall</b>	n=50	media	-0,1829660	-0,0088740	-0,1089060	0,2156887	0,3544900
		desvio	0,0357120	0,0278681	0,0423439	0,0670995	0,0720592
	n=100	media	-0,1829670	-0,0089230	-0,1086470	0,2164497	0,3541547
		desvio	0,0357872	0,0278738	0,0424704	0,0678881	0,0724756
<b>media poblacional</b>			<b>-0,1828810</b>	<b>-0,0090930</b>	<b>-0,1085770</b>	<b>0,2142594</b>	<b>0,3500000</b>
	n=10	media	-1,3175940	-0,0604920	-0,6433100	0,5912814	0,7333745
		desvio	0,3316712	0,1483913	0,3003940	0,0997550	0,0646892
	n=20	media	-1,3096730	-0,0608940	-0,6379240	0,5912894	0,7335678
		desvio	0,3358816	0,1504268	0,3017004	0,1019749	0,0664332
<b>Kuder &amp; Richardson</b>	n=50	media	-1,3095050	-0,0580140	-0,6421320	0,5925010	0,7333130
		desvio	0,3411964	0,1497476	0,3065340	0,1022390	0,0673650
	n=100	media	-1,3102910	-0,0588530	-0,6428120	0,5915239	0,7332410
		desvio	0,3429508	0,1514438	0,3077022	0,1036567	0,0674680
<b>media poblacional</b>			<b>-1,2873210</b>	<b>-0,0483430</b>	<b>-0,6106870</b>	<b>0,6036364</b>	<b>0,7368421</b>
	n=10	media	0,0171980	0,0427438	0,0918742	0,5273682	1,5716784
		desvio	0,0168728	0,0297760	0,0417529	0,2074119	0,4598429
	n=20	media	0,0173259	0,0425010	0,0925361	0,5267841	1,5852150
		desvio	0,0172561	0,0303284	0,0430835	0,2109639	0,4844709
<b>Kulczynski_1</b>	n=50	media	0,0172676	0,0423054	0,0927885	0,5280759	1,5870079
		desvio	0,0173138	0,0307582	0,0433436	0,2136590	0,4891777
	n=100	media	0,0172598	0,0427398	0,0925826	0,5282038	1,5807329
		desvio	0,0173319	0,0308446	0,0433593	0,2154544	0,4890329
<b>media poblacional</b>			<b>0,0169492</b>	<b>0,0416667</b>	<b>0,0909091</b>	<b>0,5000000</b>	<b>1,5000000</b>
	n=10	media	0,0325180	0,2214718	0,1624245	0,5958771	0,7501759
		desvio	0,0314220	0,1318768	0,0624588	0,0806548	0,0528877
	n=20	media	0,0325458	0,2209621	0,1621688	0,5952433	0,7498720
		desvio	0,0315558	0,1334578	0,0631030	0,0814694	0,0537804
<b>Kulczynski_2</b>	n=50	media	0,0328861	0,2222048	0,1623396	0,5956658	0,7496834
		desvio	0,0322063	0,1334326	0,0638369	0,0822320	0,0542229
	n=100	media	0,0327989	0,2216152	0,1626677	0,5954540	0,7500471
		desvio	0,0319562	0,1339013	0,0638518	0,0832460	0,0542775
<b>media poblacional</b>			<b>0,0327957</b>	<b>0,2212766</b>	<b>0,1625000</b>	<b>0,5952381</b>	<b>0,7500000</b>
	n=10	media	0,0161237	0,0377864	0,0760943	0,2461696	0,3737170
		desvio	0,0152932	0,0247398	0,0295019	0,0478296	0,0271030
	n=20	media	0,0160743	0,0374807	0,0757608	0,2473178	0,3734452
		desvio	0,0154164	0,0250671	0,0297955	0,0485907	0,0273440
<b>Lamont &amp; Grant</b>	n=50	media	0,0161090	0,0378037	0,0759660	0,2466786	0,3735937
		desvio	0,0157130	0,0253343	0,0301025	0,0490823	0,0273783
	n=100	media	0,0161789	0,0377981	0,0759947	0,2466891	0,3738210
		desvio	0,0157332	0,0252496	0,0302176	0,0491475	0,0273326
<b>media poblacional</b>			<b>0,0163934</b>	<b>0,0384615</b>	<b>0,0769231</b>	<b>0,2500000</b>	<b>0,3750000</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
<b>Legendre &amp; Legendre</b>	n=10	media	0,0468408	0,1076249	0,2091091	0,5895163	0,8150623
		desvio	0,0442348	0,0691361	0,0766030	0,0939607	0,0434024
	n=20	media	0,0477075	0,1070359	0,2094438	0,5888439	0,8152463
		desvio	0,0450914	0,0691652	0,0773916	0,0960220	0,0433158
	n=50	media	0,0470140	0,1081819	0,2092190	0,5900512	0,8154834
		desvio	0,0452022	0,0697507	0,0779745	0,0964976	0,0435477
	n=100	media	0,0468732	0,1074134	0,2095109	0,5903215	0,8155747
		desvio	0,0451092	0,0695052	0,0779457	0,0966633	0,0438971
	<b>media poblacional</b>		<b>0,0483871</b>	<b>0,1111111</b>	<b>0,2142857</b>	<b>0,6000000</b>	<b>0,8181818</b>
	<b>Loevinger</b>	n=10	media	-0,4557170	-0,1341050	-0,3388030	0,7688501
desvio			0,0847040	0,4552000	0,1348112	0,1475302	0,0922482
n=20		media	<b>-0,4539110</b>	-0,1333330	-0,3382350	0,7687882	<b>0,6294320</b>
		desvio	0,0858782	0,4612975	0,1372617	0,1478605	0,0933741
n=50		media	<b>-0,4547430</b>	-0,1335210	-0,3384950	0,7700611	<b>0,6294599</b>
		desvio	0,0865614	0,4673974	0,1366767	0,1497066	0,0935028
n=100		media	<b>-0,4551010</b>	-0,1361830	-0,3382390	0,7708977	<b>0,6293485</b>
		desvio	0,0872420	0,4659820	0,1381595	0,1500682	0,0941627
<b>media poblacional</b>		<b>-0,4009660</b>	<b>-0,1320750</b>	<b>-0,3333330</b>	<b>0,7685185</b>	<b>0,5833333</b>	
<b>Maxwell &amp; Pilliner</b>		n=10	media	-0,3890900	-0,0233550	-0,2318920	0,4282294
	desvio		0,0598829	0,0681422	0,0852251	0,1000284	0,0817411
	n=20	media	-0,3898490	-0,0224340	-0,2328210	0,4281902	0,5822311
		desvio	0,0609376	0,0702012	0,0856871	0,1005682	0,0815571
	n=50	media	-0,3894290	-0,0232390	-0,2330960	0,4272125	0,5822293
		desvio	0,0618343	0,0713597	0,0872724	0,1018561	0,0830949
	n=100	media	-0,3897010	-0,0232170	-0,2327400	0,4279196	0,5829334
		desvio	0,0620179	0,0711994	0,0875700	0,1018308	0,0831706
	<b>media poblacional</b>		<b>-0,3916020</b>	<b>-0,0236010</b>	<b>-0,2339180</b>	<b>0,4322917</b>	<b>0,5833333</b>
	<b>McConnaughey</b>	n=10	media	-0,9342390	-0,5552480	-0,6761340	0,1917487
desvio			0,0618089	0,2649226	0,1239524	0,1600879	0,1058904
n=20		media	-0,9343410	-0,5573530	-0,6749460	0,1905663	0,5020845
		desvio	0,0629713	0,2638230	0,1267530	0,1634995	0,1080262
n=50		media	-0,9341580	-0,5578620	-0,6747530	0,1898496	0,5005338
		desvio	0,0642631	0,2668639	0,1274873	0,1654351	0,1082721
n=100		media	-0,9343890	-0,5569140	-0,6752840	0,1909130	0,4993605
		desvio	0,0639467	0,2677614	0,1277997	0,1666513	0,1091637
<b>media poblacional</b>		<b>-0,9344090</b>	<b>-0,5574470</b>	<b>-0,6750000</b>	<b>0,1904762</b>	<b>0,5000000</b>	
<b>Michael</b>		n=10	media	-0,6284660	-0,0281560	-0,3853840	0,3830654
	desvio		0,0983932	0,0836259	0,1412607	0,1028124	0,0744263
	n=20	media	-0,6264350	-0,0279860	-0,3848610	0,3829404	0,8082625
		desvio	0,0998977	0,0835421	0,1410131	0,1045721	0,0753476
	n=50	media	-0,6276170	-0,0276940	-0,3864440	0,3831345	0,8074335
		desvio	0,1007857	0,0844610	0,1427343	0,1059451	0,0766335
	n=100	media	-0,6272440	-0,0274460	-0,3859800	0,3829836	<b>0,8075394</b>
		desvio	0,1012329	0,0844738	0,1430670	0,1057889	0,0766922
	<b>media poblacional</b>		<b>-0,6431620</b>	<b>-0,0279550</b>	<b>-0,3960400</b>	<b>0,3905882</b>	<b>0,8235294</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
<b>Mountford</b>	n=10	media	0,0011263	0,0124362	0,0061900	0,0793600	0,0799462
		desvio	0,0011032	0,0094385	0,0027099	0,0255325	0,0217971
	n=20	media	0,0011310	0,0126102	0,0061416	0,0798124	0,0800485
		desvio	0,0011149	0,0098394	0,0027082	0,0266917	0,0227020
	n=50	media	0,0011405	0,0126699	0,0061491	0,0797918	0,0801138
		desvio	0,0011412	0,0100287	0,0027584	0,0269630	0,0229709
	n=100	media	0,0011433	0,0125993	0,0061458	<b>0,0796108</b>	<b>0,0802227</b>
		desvio	0,0011545	0,0100260	0,0027640	0,0271269	0,0233035
	<b>media poblacional</b>		<b>0,0011117</b>	<b>0,0109290</b>	<b>0,0059701</b>	<b>0,0735294</b>	<b>0,0750000</b>
	<b>Ochiai</b>	n=10	media	0,0326172	0,1277485	0,1570400	0,5391731
desvio			0,0311665	0,0742887	0,0606573	0,0872838	0,0524120
n=20		media	0,0325968	0,1272218	0,1566005	0,5410475	0,7492458
		desvio	0,0314968	0,0751020	0,0612268	0,0877482	0,0539202
n=50		media	0,0326753	0,1267111	0,1568714	0,5414950	0,7488630
		desvio	0,0317470	0,0757055	0,0615836	0,0891312	0,0541931
n=100		media	0,0326187	0,1274754	0,1568367	0,5406289	0,7490016
		desvio	0,0316694	0,0760667	0,0619682	0,0898859	0,0546515
<b>media poblacional</b>		<b>0,0327913</b>	<b>0,1304656</b>	<b>0,1581139</b>	<b>0,5455447</b>	<b>0,7500000</b>	
<b>Pearson &amp; Heron</b>		n=10	media	<b>-0,2303160</b>	<b>-0,0405400</b>	<b>0,0045812</b>	0,1316896
	desvio		0,6143721	0,7061737	0,6969421	0,7383677	0,7011335
	n=20	media	<b>-0,2204170</b>	<b>-0,0453240</b>	<b>-0,0015210</b>	<b>0,1355391</b>	<b>0,0097168</b>
		desvio	0,6305997	0,7030412	0,7038064	0,7384690	0,7032724
	n=50	media	<b>-0,2140780</b>	<b>-0,0325080</b>	<b>0,0017593</b>	<b>0,1290039</b>	<b>-0,0066980</b>
		desvio	0,6309452	0,7078911	0,7067877	0,7389044	0,7056174
	n=100	media	<b>-0,2181450</b>	<b>-0,0347330</b>	<b>-0,0040940</b>	<b>0,1293303</b>	<b>0,0002703</b>
		desvio	0,6324070	0,7085121	0,7071879	0,7404395	0,7076935
	<b>media poblacional</b>		<b>-0,8452610</b>	<b>-0,7962900</b>	<b>-0,4876450</b>	<b>-0,2835990</b>	<b>0,7243160</b>
	<b>Pearson_1</b>	n=10	media	15,6313860	<b>1,0733598</b>	6,3027759	21,4097490
desvio			4,6888316	1,2071510	4,0197646	8,5740276	9,5229778
n=20		media	15,6219660	<b>1,0939047</b>	6,3485667	21,4666440	34,6432190
		desvio	4,7196655	1,2638625	4,0691418	8,7352878	9,5579403
n=50		media	15,6408720	<b>1,0805770</b>	6,3091202	21,4742940	34,7379710
		desvio	4,8066380	1,2792869	4,1141739	8,7669127	9,5602958
n=100		media	15,6168220	<b>1,0752395</b>	6,2994614	21,4456010	34,7289890
		desvio	4,7823991	1,2839966	4,1329899	8,7617687	9,6581699
<b>media poblacional</b>		<b>15,3364950</b>	<b>0,1035306</b>	<b>5,5555556</b>	<b>20,7100770</b>	<b>34,0277780</b>	
<b>Pearson_2</b>		n=10	media	0,3613577	<b>0,0849731</b>	0,2287645	0,4075250
	desvio		0,0485786	0,0574958	0,0780337	0,0739990	0,0529270
	n=20	media	0,3623646	<b>0,0852434</b>	0,2261596	0,4076979	0,5000247
		desvio	0,0492386	0,0581137	0,0789349	0,0746091	0,0543400
	n=50	media	0,3618268	<b>0,0840034</b>	0,2269575	0,4086245	0,5013746
		desvio	0,0500427	0,0581317	0,0799764	0,0757809	0,0540824
	n=100	media	0,3621488	<b>0,0840625</b>	0,2266736	0,4079478	0,5008877
		desvio	0,0501043	0,0582055	0,0797854	0,0758934	0,0542521
	<b>media poblacional</b>		<b>0,3646529</b>	<b>0,0321595</b>	<b>0,2294157</b>	<b>0,4142086</b>	<b>0,5038710</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	0,3017601	0,0546340	0,2747993	0,0411079	0,4013718
		desvio	0,0303740	0,0285868	0,0377791	0,0276895	0,0959583
	n=20	media	0,3015547	0,0551616	0,2744019	0,0413033	0,4024644
		desvio	0,0306715	0,0291626	0,0384199	0,0282453	0,0987214
<b>Peirce_1</b>	n=50	media	0,3017197	0,0547339	0,2743391	0,0414793	0,4025295
		desvio	0,0308360	0,0292137	0,0385740	0,0282867	0,0987127
	n=100	media	0,3014872	0,0550767	0,2747554	0,0415006	0,4032961
		desvio	0,0309413	0,0292734	0,0385893	0,0286262	0,0994668
<b>media poblacional</b>			<b>0,3027956</b>	<b>0,0558195</b>	<b>0,2758621</b>	<b>0,0414201</b>	<b>0,4000000</b>
	n=10	media	-0,3878590	-0,0136140	-0,2079790	0,3280880	0,5829920
		desvio	0,0664567	0,0424614	0,0778329	0,0921034	0,0822603
	n=20	media	-0,3875320	-0,0137430	-0,2089600	0,3289857	0,5823890
		desvio	0,0675365	0,0433431	0,0796949	0,0926582	0,0830604
<b>Peirce_2</b>	n=50	media	-0,3878970	-0,0139420	-0,2084340	0,3288157	0,5836131
		desvio	0,0671717	0,0431986	0,0803273	0,0937519	0,0832165
	n=100	media	-0,3879810	-0,0140640	-0,2081760	0,3292598	0,5831144
		desvio	0,0676954	0,0432097	0,0804485	0,0937470	0,0838463
<b>media poblacional</b>			<b>-0,3880320</b>	<b>-0,0140510</b>	<b>-0,2083330</b>	<b>0,3293651</b>	<b>0,5833333</b>
	n=10	media	-0,3938380	-0,0702030	-0,2657930	0,6288927	0,5829695
		desvio	0,0662683	0,2477370	0,0970709	0,1163015	0,0821459
	n=20	media	-0,3947670	-0,0714050	-0,2668170	0,6293709	0,5846547
		desvio	0,0674217	0,2492406	0,0983610	0,1196369	0,0830092
<b>Peirce_3</b>	n=50	media	-0,3952300	-0,0751560	-0,2665930	0,6292853	0,5833523
		desvio	0,0674394	0,2506732	0,0992316	0,1198798	0,0835852
	n=100	media	-0,3953670	-0,0739220	-0,2668130	0,6288465	0,5833050
		desvio	0,0679976	0,2518541	0,0996565	0,1203636	0,0839461
<b>media poblacional</b>			<b>-0,3952380</b>	<b>-0,0736840</b>	<b>-0,2666670</b>	<b>0,6287879</b>	<b>0,5833333</b>
	n=10	media	0,2592365	0,3518206	0,2915148	0,6680567	0,6687394
		desvio	0,0380226	0,0440746	0,0405872	0,0534518	0,0542707
	n=20	media	0,2595176	0,3532548	0,2917861	0,6690347	0,6687392
		desvio	0,0389728	0,0448613	0,0412158	0,0552401	0,0546046
<b>Roger &amp; Tanimoto</b>	n=50	media	0,2592167	0,3533069	0,2919786	0,6686121	0,6685166
		desvio	0,0389426	0,0456859	0,0414495	0,0552088	0,0553871
	n=100	media	0,2591522	0,3527731	0,2913863	0,6684721	0,6683558
		desvio	0,0391367	0,0456721	0,0414477	0,0554712	0,0553787
<b>media poblacional</b>			<b>0,2578616</b>	<b>0,3513514</b>	<b>0,2903226</b>	<b>0,6666667</b>	<b>0,6666667</b>
	n=10	media	0,3028988	0,3749743	0,3711807	0,6834288	0,7902374
		desvio	0,0296146	0,0345185	0,0419712	0,0552417	0,0413445
	n=20	media	0,3029638	0,3751414	0,3713193	0,6836683	0,7898787
		desvio	0,0303406	0,0356137	0,0419654	0,0567649	0,0415719
<b>Rogot &amp; Goldberg</b>	n=50	media	0,3031102	0,3749421	0,3718422	0,6837291	0,7903511
		desvio	0,0304311	0,0359289	0,0425978	0,0567211	0,0416902
	n=100	media	0,3029023	0,3749397	0,3712698	0,6837845	0,7897588
		desvio	0,0305889	0,0358899	0,0426420	0,0574441	0,0418512
<b>media poblacional</b>			<b>0,3041632</b>	<b>0,3762994</b>	<b>0,3732194</b>	<b>0,6875000</b>	<b>0,7916667</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	0,0100210	0,0199430	0,0500030	0,0996970	0,2997300
		desvio	0,0095029	0,0133272	0,0211271	0,0290092	0,0447369
	n=20	media	0,0100350	0,0198780	0,0497835	0,1000415	0,2999815
		desvio	0,0097684	0,0136848	0,0215914	0,0294411	0,0454539
<b>Russell &amp; Rao</b>	n=50	media	0,0100352	0,0201044	0,0498988	0,0999812	0,3001744
		desvio	0,0098633	0,0139323	0,0215023	0,0297671	0,0456183
	n=100	media	0,0099906	0,0199633	0,0500521	0,1000517	0,3001019
		desvio	0,0099144	0,0139578	0,0217721	0,0300181	0,0456797
<b>media poblacional</b>			<b>0,0100000</b>	<b>0,0200000</b>	<b>0,0500000</b>	<b>0,1000000</b>	<b>0,3000000</b>
	n=10	media	-0,3930810	-0,2511650	-0,2568010	0,3667662	0,5798584
		desvio	0,0601414	0,0703064	0,0827098	0,1118686	0,0821378
	n=20	media	-0,3943120	-0,2502050	-0,2571710	0,3681942	0,5797654
		desvio	0,0605171	0,0707972	0,0843249	0,1139351	0,0818887
<b>Scott</b>	n=50	media	-0,3939590	-0,2505030	-0,2563840	0,3684488	0,5798696
		desvio	0,0611208	0,0715501	0,0850716	0,1139588	0,0829808
	n=100	media	-0,3938350	-0,2499270	-0,2577420	0,3677377	0,5805520
		desvio	0,0612653	0,0720827	0,0851719	0,1146693	0,0833198
<b>media poblacional</b>			<b>-0,3916740</b>	<b>-0,2474010</b>	<b>-0,2535610</b>	<b>0,3750000</b>	<b>0,5833333</b>
	n=10	media	0,0358440	0,3982330	0,1983911	0,8327023	0,7838393
		desvio	0,0340260	0,2390407	0,0791367	0,1091490	0,0583316
	n=20	media	0,0361079	0,4012354	0,1986656	0,8346419	<b>0,7834679</b>
		desvio	0,0347529	0,2432973	0,0806542	0,1092154	0,0592091
<b>Simpson</b>	n=50	media	0,0358465	0,4023006	0,2005194	0,8339785	<b>0,7833880</b>
		desvio	0,0350836	0,2468636	0,0810079	0,1111846	0,0595983
	n=100	media	0,0360703	0,4000764	0,2003810	0,8334845	<b>0,7831361</b>
		desvio	0,0352955	0,2472269	0,0808684	0,1118109	0,0596966
<b>media poblacional</b>			<b>0,0333333</b>	<b>0,4000000</b>	<b>0,2000000</b>	<b>0,8333333</b>	<b>0,7500000</b>
	n=10	media	0,4095270	0,5197980	0,4502560	0,8005050	0,7996840
		desvio	0,0481420	0,0480197	0,0486670	0,0391080	0,0388542
	n=20	media	0,4104145	0,5190225	0,4495845	0,7999560	0,7999890
		desvio	0,0486986	0,0492611	0,0488476	0,0397568	0,0392747
<b>Sokal &amp; Michener</b>	n=50	media	0,4102018	0,5198750	0,4498634	0,8002062	0,8001046
		desvio	0,0489143	0,0497834	0,0495711	0,0399507	0,0398683
	n=100	media	0,4097383	0,5200829	0,4501175	0,7999150	0,7999851
		desvio	0,0491027	0,0497331	0,0497622	0,0398706	0,0397751
<b>media poblacional</b>			<b>0,4100000</b>	<b>0,5200000</b>	<b>0,4500000</b>	<b>0,8000000</b>	<b>0,8000000</b>
	n=10	media	0,0084535	0,0204986	0,0436917	0,2027172	0,4312343
		desvio	0,0082291	0,0140130	0,0191854	0,0624154	0,0694920
	n=20	media	0,0084505	0,0207729	0,0438659	0,2028112	0,4312888
		desvio	0,0083237	0,0144112	0,0192966	0,0627600	0,0704688
<b>Sokal &amp; Sneath_1</b>	n=50	media	0,0085082	0,0206147	0,0439425	0,2029696	0,4321266
		desvio	0,0084508	0,0144376	0,0196126	0,0634904	0,0715305
	n=100	media	0,0084772	0,0207429	0,0439500	0,2034525	0,4322168
		desvio	0,0084496	0,0146176	0,0195888	0,0638696	0,0711644
<b>media poblacional</b>			<b>0,0084034</b>	<b>0,0204082</b>	<b>0,0434783</b>	<b>0,2000000</b>	<b>0,4285714</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	0,3041172	0,4779859	0,3814908	0,7396224	0,7917929
		desvio	0,0301927	0,0711323	0,0428318	0,0469112	0,0397633
	n=20	media	0,3042904	0,4783306	0,3811957	0,7398875	0,7922709
		desvio	0,0306608	0,0722965	0,0436659	0,0478787	0,0407902
<b>Sokal &amp; Sneath_2</b>	n=50	media	0,3040948	0,4785269	0,3811078	0,7394264	0,7916427
		desvio	0,0306161	0,0724032	0,0442101	0,0485982	0,0415189
	n=100	media	0,3042555	0,4781499	0,3812635	0,7395583	0,7916583
		desvio	0,0307448	0,0729675	0,0442558	0,0484477	0,0413808
<b>media poblacional</b>			<b>0,3041825</b>	<b>0,4780663</b>	<b>0,3812500</b>	<b>0,7395382</b>	<b>0,7916667</b>
	n=10	media	0,7052639	1,1048154	0,8329275	4,2175535	4,2239838
		desvio	0,1408621	0,2191841	0,1651262	1,1158882	1,0794446
	n=20	media	0,7056651	1,1070911	0,8332877	4,2247081	4,2019289
		desvio	0,1424950	0,2214410	0,1664325	1,1178183	1,1205892
<b>Sokal &amp; Sneath_3</b>	n=50	media	0,7073828	1,1087986	0,8332482	4,2268072	4,2163437
		desvio	0,1450701	0,2256525	0,1690452	1,1434649	1,1563972
	n=100	media	0,7069842	1,1055699	0,8346531	4,2230746	4,2191282
		desvio	0,1453581	0,2243138	0,1711356	1,1602089	1,1524565
<b>media poblacional</b>			<b>0,6949153</b>	<b>1,0833333</b>	<b>0,8181818</b>	<b>4,0000000</b>	<b>4,0000000</b>
	n=10	media	0,5802633	0,6831695	0,6186101	0,8880152	0,8879973
		desvio	0,0489298	0,0419820	0,0461700	0,0242299	0,0240541
	n=20	media	0,5793986	0,6828868	0,6190964	0,8885110	0,8882635
		desvio	0,0488617	0,0429289	0,0467775	0,0243874	0,0242928
<b>Sokal &amp; Sneath_4</b>	n=50	media	0,5800337	0,6829228	0,6188998	0,8882251	0,8883128
		desvio	0,0492113	0,0430312	0,0471036	0,0246126	0,0245287
	n=100	media	0,5797824	0,6827597	0,6192745	0,8883221	0,8883039
		desvio	0,0495233	0,0433951	0,0474062	0,0246545	0,0247132
<b>media poblacional</b>			<b>0,5815603</b>	<b>0,6842105</b>	<b>0,6206897</b>	<b>0,8888889</b>	<b>0,8888889</b>
	n=10	media	0,0191676	0,0893654	0,0936197	0,4777529	0,6250143
		desvio	0,0182991	0,0535553	0,0384320	0,0840608	0,0654347
	n=20	media	0,0189830	0,0900044	0,0939428	0,4768193	0,6244723
		desvio	0,0186098	0,0546162	0,0388142	0,0857728	0,0662030
<b>Sokal &amp; Sneath_5</b>	n=50	media	0,0188401	0,0908652	0,0936790	0,4764255	0,6257032
		desvio	0,0185829	0,0551332	0,0394352	0,0869121	0,0663306
	n=100	media	0,0188050	0,0899451	0,0938886	0,4770587	0,6247284
		desvio	0,0187959	0,0548577	0,0395139	0,0868605	0,0662603
<b>media poblacional</b>			<b>0,0188732</b>	<b>0,0919319</b>	<b>0,0942809</b>	<b>0,4797564</b>	<b>0,6250000</b>
	n=10	media	0,0322692	0,0768160	0,1513520	0,4926520	0,7475530
		desvio	0,0306437	0,0490810	0,0589860	0,0960990	0,0532220
	n=20	media	0,0320070	0,0754370	0,1524280	0,4946570	0,7480080
		desvio	0,0306888	0,0496190	0,0604050	0,0977690	0,0540820
<b>Sorensen</b>	n=50	media	0,0325234	0,0756380	0,1518770	0,4942200	0,7478520
		desvio	0,0315929	0,0505730	0,0601690	0,0980150	0,0543300
	n=100	media	0,0323844	0,0753070	0,1518530	0,4931300	0,7477870
		desvio	0,0315350	0,0503950	0,0604650	0,0984170	0,0547830
<b>media poblacional</b>			<b>0,0327869</b>	<b>0,0769231</b>	<b>0,1538462</b>	<b>0,5000000</b>	<b>0,7500000</b>

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30	
		b	0,29	0,03	0,20	0,02	0,10	
		c	0,30	0,45	0,35	0,18	0,10	
		d	0,40	0,50	0,40	0,70	0,50	
Sorgenfrei	n=10	media	0,0020815	0,0217698	0,0283828	0,3006761	0,5637490	
		desvio	0,0028157	0,0202964	0,0199519	0,0916518	0,0784753	
	n=20	media	0,0020676	0,0222862	0,0285846	0,3006057	0,5637141	
		desvio	0,0030022	0,0211220	0,0203213	0,0941300	0,0802626	
	n=50	media	<b>0,0020745</b>	0,0219422	0,0285185	0,3010490	0,5645189	
		desvio	0,0031560	0,0208936	0,0206118	0,0945294	0,0810332	
	n=100	media	<b>0,0020740</b>	<b>0,0220024</b>	0,0284539	0,3008383	0,5634583	
		desvio	0,0031910	0,0211909	0,0206977	0,0946495	0,0811833	
	<b>media poblacional</b>		<b>0,0010753</b>	<b>0,0170213</b>	<b>0,0250000</b>	<b>0,2976190</b>	<b>0,5625000</b>	
	Stiles	n=10	media	2,5714354	<b>-2,1129150</b>	1,2536200	2,7822402	3,4288566
			desvio	0,3436653	2,0261814	1,1000750	0,5350932	0,3003366
		n=20	media	2,5658430	<b>-2,0700500</b>	1,2413889	2,7819390	3,4336771
desvio			0,3543644	2,0615635	1,1717564	0,5556334	0,3042858	
n=50		media	2,5660697	<b>-2,0857730</b>	1,2432759	2,7832721	3,4307291	
		desvio	0,3589271	2,0675639	1,2031816	0,5720630	0,3080383	
n=100		media	2,5677912	<b>-2,0903540</b>	<b>1,2356160</b>	2,7823697	3,4295006	
		desvio	0,3607072	2,0767672	1,2455735	0,5743053	0,3078755	
<b>media poblacional</b>		<b>2,6059717</b>	<b>-3,9624840</b>	<b>1,5040774</b>	<b>2,8740459</b>	<b>3,4544419</b>		
Tarwid		n=10	media	-0,8220650	-0,1816280	-0,3573140	0,4955386	0,3051103
			desvio	0,1576799	0,3588570	0,1695296	0,0670817	0,0465425
		n=20	media	-0,8230670	-0,1782130	-0,3558780	0,4948155	0,3048932
	desvio		0,1590008	0,3676348	0,1713375	0,0684553	0,0467374	
	n=50	media	-0,8217430	-0,1800310	-0,3574740	0,4951714	0,3053085	
		desvio	0,1612947	0,3720257	0,1733464	0,0687766	0,0474235	
	n=100	media	-0,8226100	<b>-0,1823560</b>	-0,3579080	0,4950403	0,3050458	
		desvio	0,1611080	0,3749978	0,1748687	0,0691466	0,0475456	
	<b>media poblacional</b>		<b>-0,8058250</b>	<b>-0,0804600</b>	<b>-0,3333330</b>	<b>0,4970060</b>	<b>0,3043478</b>	
	Tschuproff	n=10	media	0,3909219	<b>0,0852845</b>	0,2357211	0,4516382	0,5838971
			desvio	0,0605870	0,0582998	0,0853044	0,0951108	0,0810274
		n=20	media	0,3903152	<b>0,0852426</b>	0,2354669	0,4520426	0,5845601
desvio			0,0604004	0,0590024	0,0867024	0,0979531	0,0827711	
n=50		media	0,3906737	<b>0,0857654</b>	0,2358681	0,4522898	0,5828161	
		desvio	0,0615447	0,0594915	0,0866105	0,0986388	0,0829943	
n=100		media	0,3902743	<b>0,0850196</b>	0,2355477	0,4526401	0,5832647	
		desvio	0,0617666	0,0593660	0,0868318	0,0988165	0,0834555	
<b>media poblacional</b>		<b>0,3916184</b>	<b>0,0321762</b>	<b>0,2357023</b>	<b>0,4550833</b>	<b>0,5833333</b>		
Yule_1		n=10	media	-0,3903120	-0,0320350	-0,2342970	0,4530797	0,5820691
			desvio	0,0603186	0,0960729	0,0854044	0,0969777	0,0816598
		n=20	media	-0,3900370	-0,0309340	-0,2352130	0,4537958	0,5841890
	desvio		0,0615810	0,0980608	0,0871624	0,0970482	0,0821907	
	n=50	media	-0,3906610	-0,0311690	-0,2344350	0,4518272	0,5835681	
		desvio	0,0615596	0,0986428	0,0875195	0,0984453	0,0827656	
	n=100	media	-0,3906070	-0,0317990	-0,2353070	0,4522309	0,5832422	
		desvio	0,0616311	0,0987277	0,0878996	0,0991170	0,0829377	
	<b>media poblacional</b>		<b>-0,3916180</b>	<b>-0,0321760</b>	<b>-0,2357020</b>	<b>0,4550833</b>	<b>0,5833333</b>	

Índices de similitud	proporciones en la población	a	0,01	0,02	0,05	0,10	0,30
		b	0,29	0,03	0,20	0,02	0,10
		c	0,30	0,45	0,35	0,18	0,10
		d	0,40	0,50	0,40	0,70	0,50
	n=10	media	-0,9097170	-0,1563220	-0,5550320	0,8936207	0,8710070
		desvio	0,0882780	0,4978581	0,1843937	0,0844430	0,0592580
	n=20	media	-0,9107140	-0,1575360	-0,5513390	0,8930072	0,8707560
		desvio	0,0902860	0,5046178	0,1898349	0,0882515	0,0608680
<b>Yule_2</b>	n=50	media	-0,9112570	-0,1622520	-0,5530090	0,8932960	0,8708490
		desvio	0,0910327	0,5070546	0,1926846	0,0886740	0,0616500
	n=100	media	-0,9104080	-0,1599230	-0,5523100	0,8936420	0,8703730
		desvio	0,0916520	0,5063326	0,1918384	0,0885430	0,0623870
<b>media poblacional</b>			<b>-0,9120880</b>	<b>-0,1489360</b>	<b>-0,5555560</b>	<b>0,9021739</b>	<b>0,8750000</b>
	n=10	media	-0,7370750	-0,1284730	-0,3180460	0,6662410	0,5961076
		desvio	0,2122640	0,4115640	0,1367010	0,1610540	0,0825331
	n=20	media	-0,7363850	-0,1241390	-0,3198660	0,6644286	0,5966624
		desvio	0,2122430	0,4233970	0,1374900	0,1613571	0,0830732
<b>Yule_3</b>	n=50	media	<b>-0,7360770</b>	-0,1270600	-0,3193910	0,6649187	0,5956701
		desvio	0,2133210	0,4274490	0,1398940	0,1646369	0,0838130
	n=100	media	<b>-0,7357850</b>	-0,1282890	-0,3173090	<b>0,6651023</b>	0,5958689
		desvio	0,2140100	0,4318750	0,1395230	0,1641930	0,0840634
<b>media poblacional</b>			<b>-0,6468730</b>	<b>-0,0748860</b>	<b>-0,3033370</b>	<b>0,6302859</b>	<b>0,5895738</b>

## ANEXO 6

Se presenta en este Anexo el incumplimiento de la Propiedad Métrica (13) por parte de las tres Disimilitudes derivadas del índice de Similitud de Sokal & Sneath\_2 con un contra ejemplo presentado por Gower & Legendre (1986). La fórmula del índice es:

$$A_{66}(j, k)^{(a,b,c,d)} = \frac{1}{4} \left( \frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{b+d} \right)$$

El índice varía entre 0 y 1, las disimilitudes se definen como se observa en la página 34.....

Se comparan tres unidades muestrales 1, 2 y 3, que contienen la presencia-ausencia de veintitrés variables.

	1	2	3
1	0	1	1
2	0	1	1
3	0	1	1
4	0	1	1
5	0	1	1
6	0	1	1
7	0	1	1
8	0	1	1
9	0	1	1
10	0	1	1
11	0	1	1
12	0	1	1
13	1	0	1
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0

Se comparan de a dos las tres unidades muestrales utilizando la tabla de contingencias de 2x2:

		Obs. 1		
		presente	ausente	
Obs. 2	presente	0	12	12
	ausente	1	10	11
		1	22	n=23

$$D_{166}(1,2)=0,6590$$

$$D_{266}(1,2)=0,8118$$

$$D_{366}(1,2)=1,1481$$

		Obs. 2		
		presente	ausente	
Obs. 3	presente	12	1	13
	ausente	0	10	10
		12	11	n=23

$$D_{166}(2,3)=0,0419$$

$$D_{266}(2,3)=0,2048$$

$$D_{366}(2,3)=0,2896$$

		Obs. 1		
		presente	ausente	
Obs. 3	presente	1	12	13
	ausente	0	10	10
		1	22	n=23

$$D_{166}(1,3)=0,3671$$

$$D_{266}(1,3)=0,6059$$

$$D_{366}(1,3)=0,8568$$

Si probamos la desigualdad triangular con la Disimilitud 1 vemos que no se cumple:

$$D_{166}(1,2) > D_{166}(1,3) + D_{166}(2,3)$$

$$0,6590 > 0,0419 + 0,3671 = 0,4090$$

Lo mismo ocurre con la Disimilitud 2:

$$D_{266}(1,2) > D_{266}(1,3) + D_{266}(2,3)$$

$$0,8118 > 0,2048 + 0,6069 = 0,8107$$

y con la Disimilitud 3:

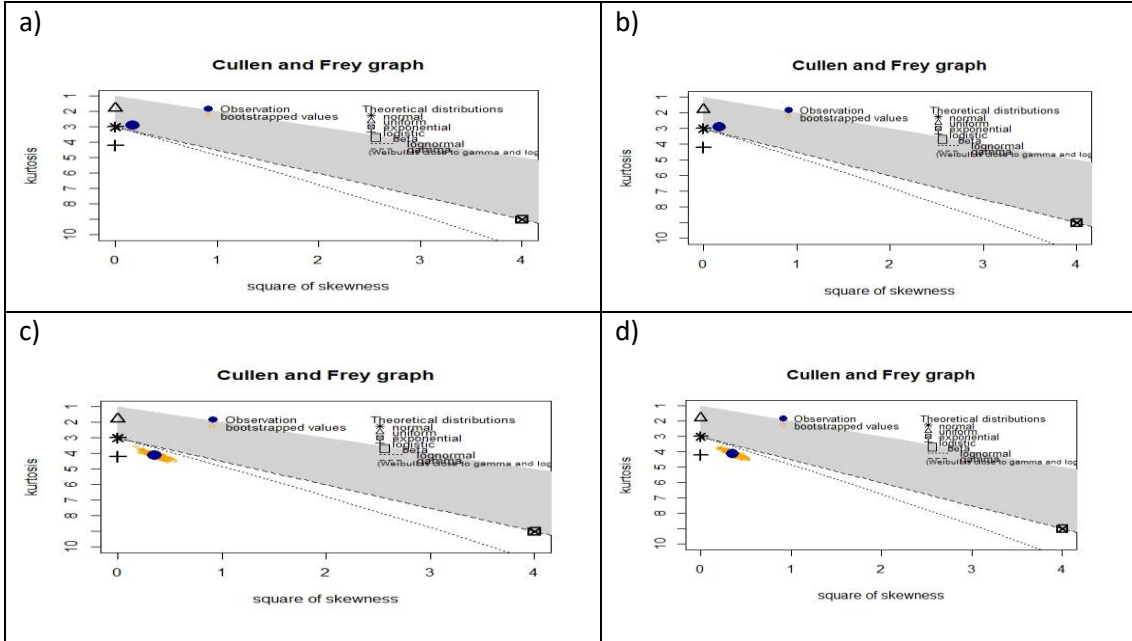
$$D_{366}(1,2) > D_{366}(1,3) + D_{366}(2,3)$$

$$1,1481 > 0,2896 + 0,8568 = 1,1464$$

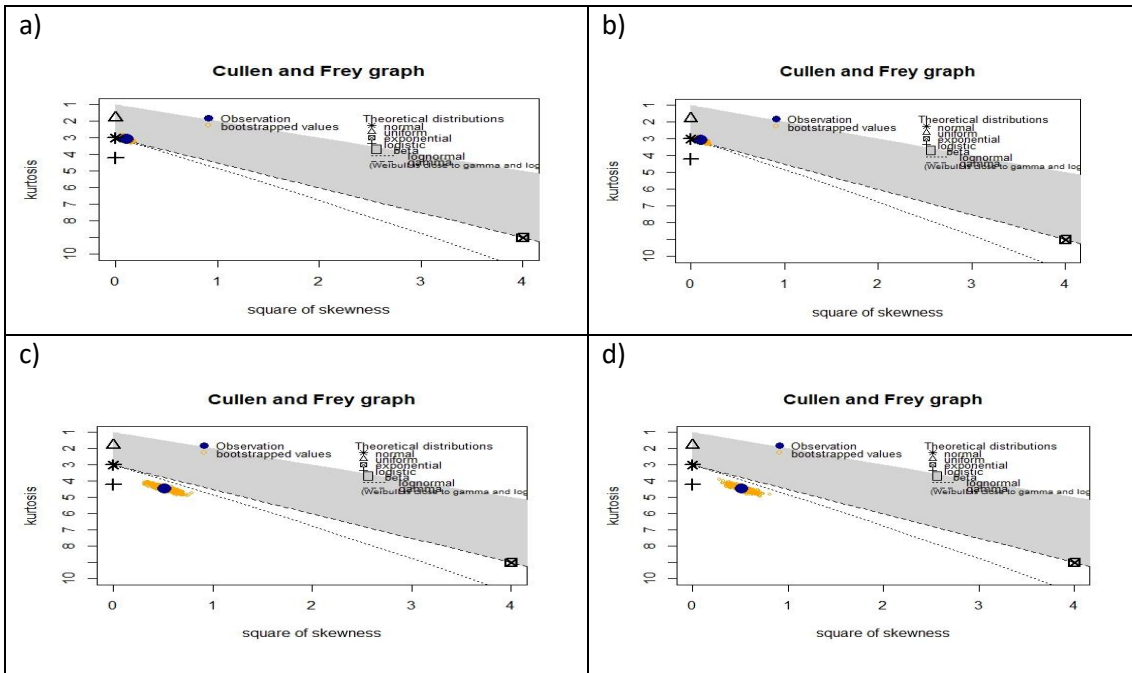


## ANEXO 7

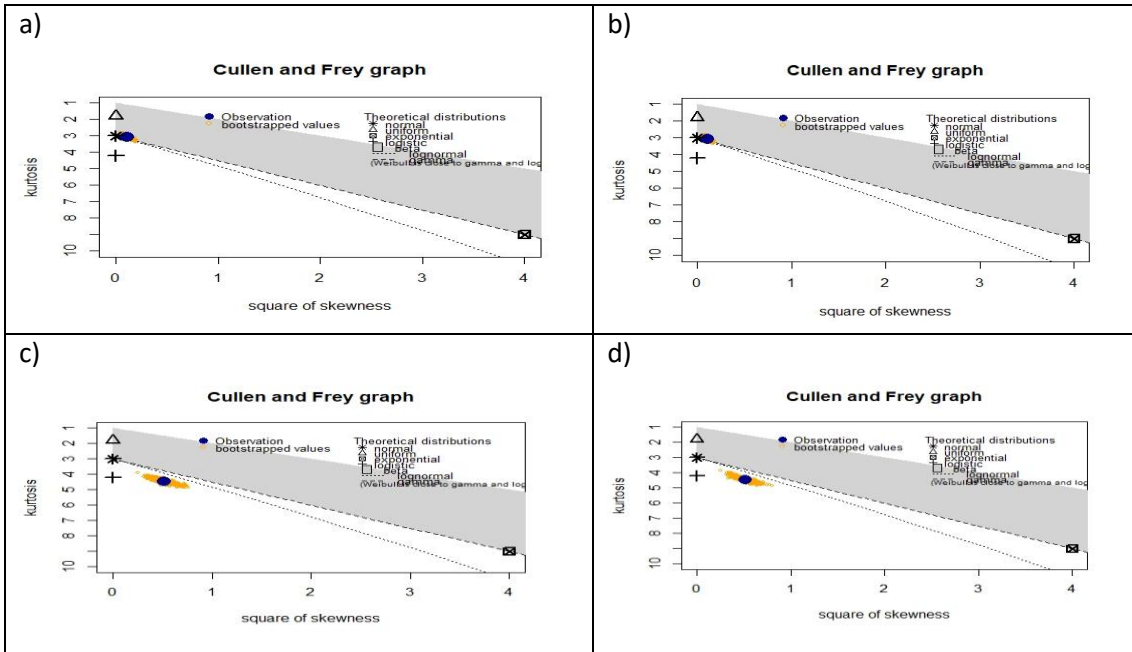
### Anónimo\_1



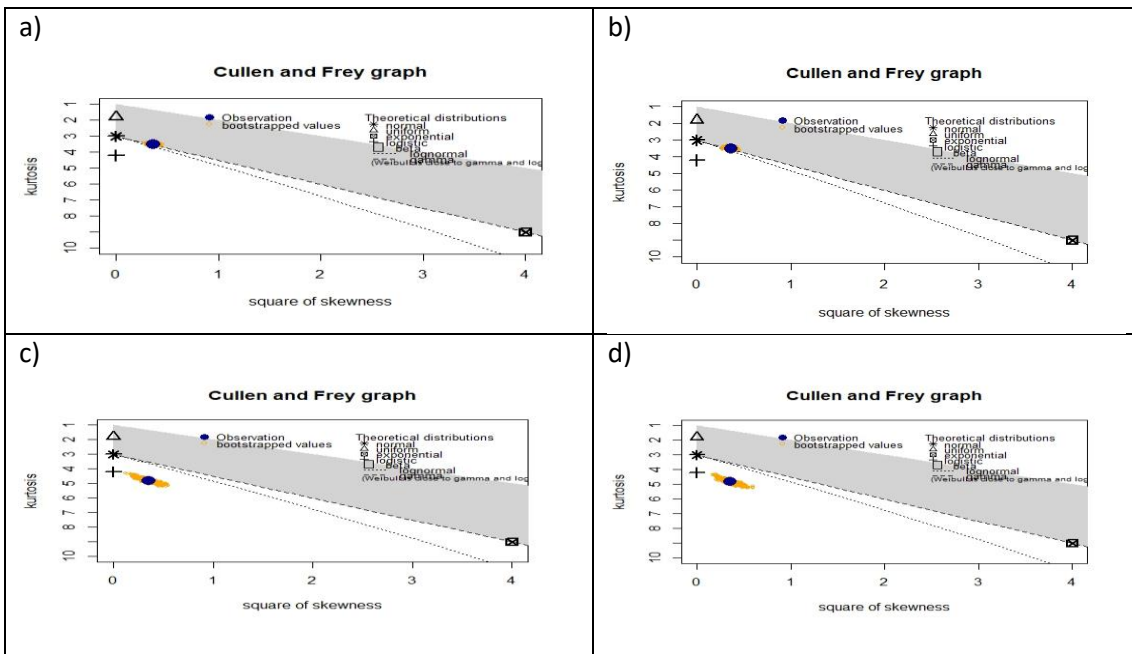
### Anónimo\_2



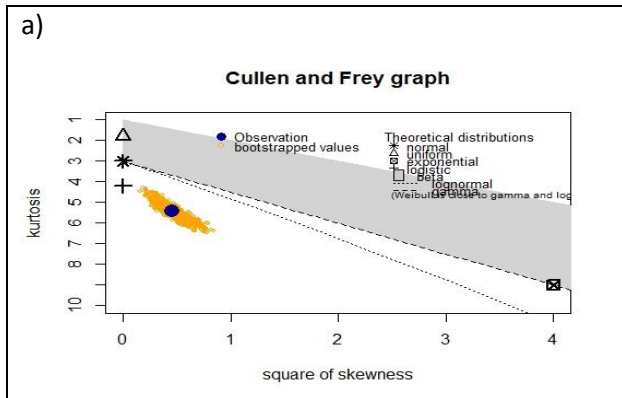
Anónimo\_3



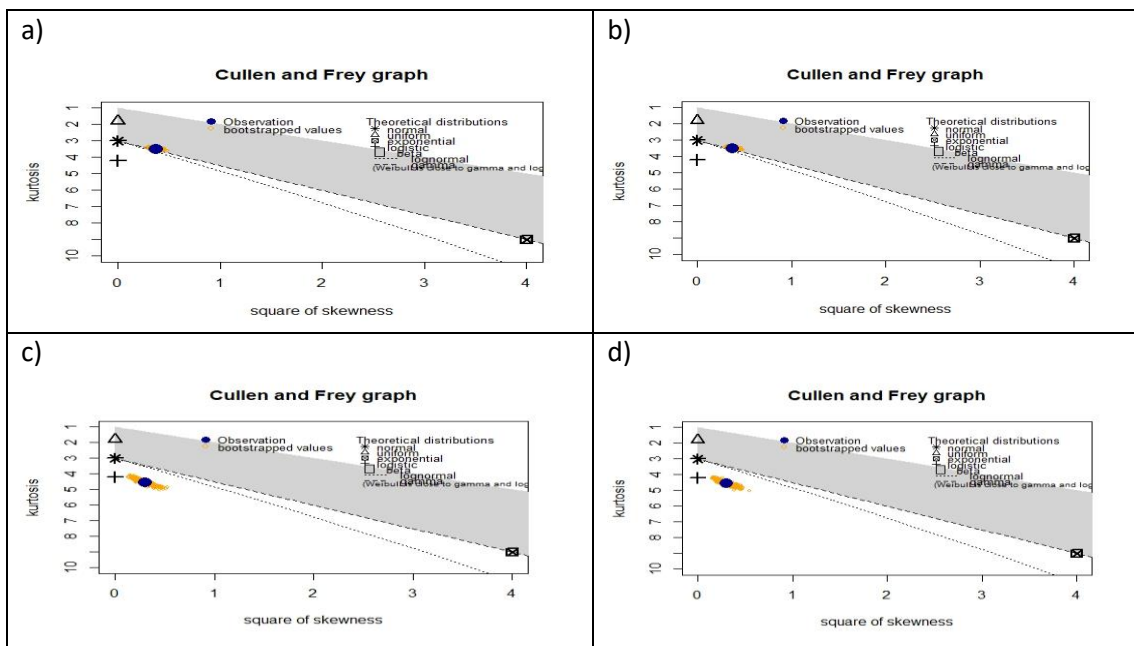
Anónimo\_4



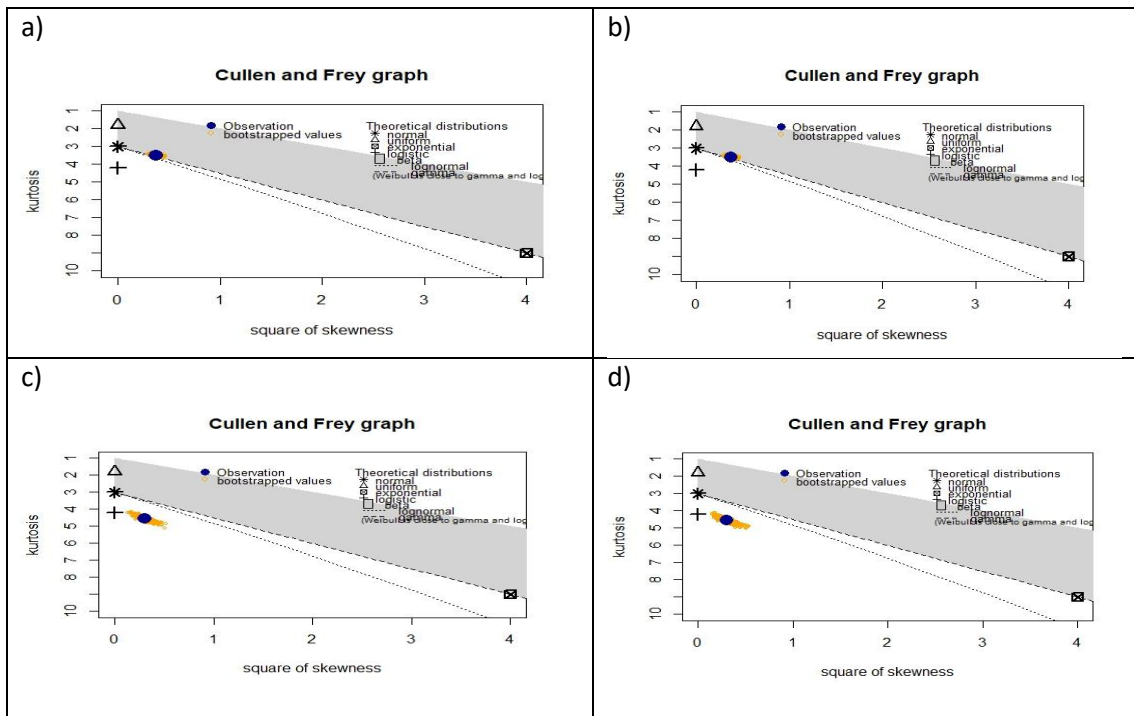
Anónimo\_5



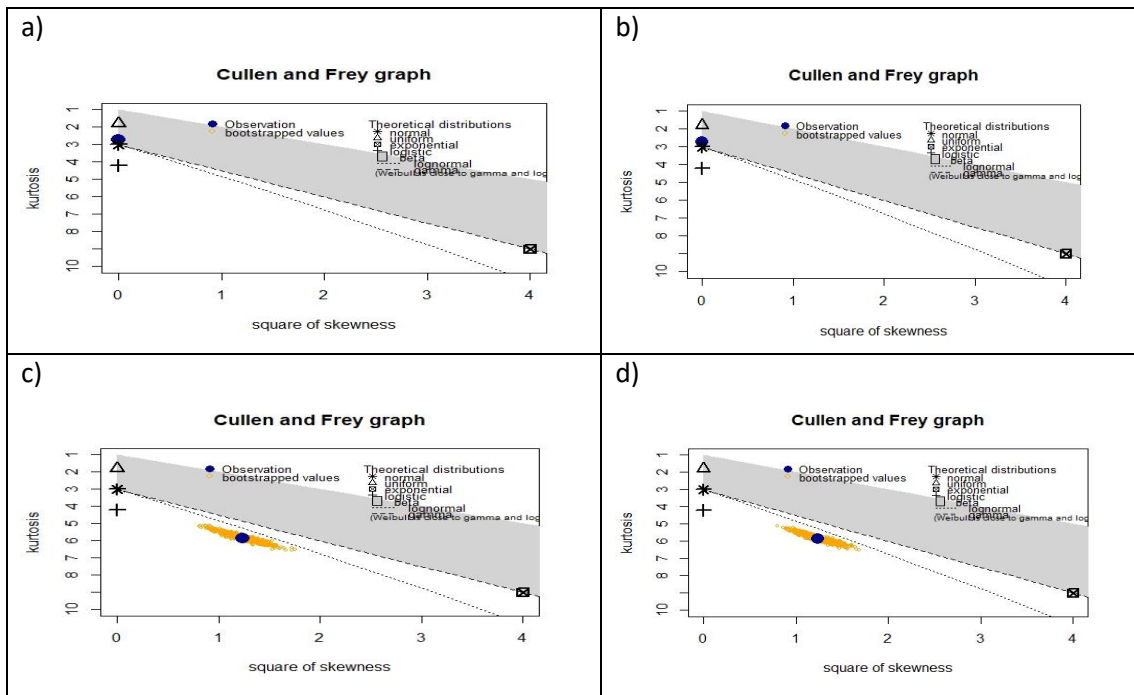
Baroni-Urbani & Buser\_1



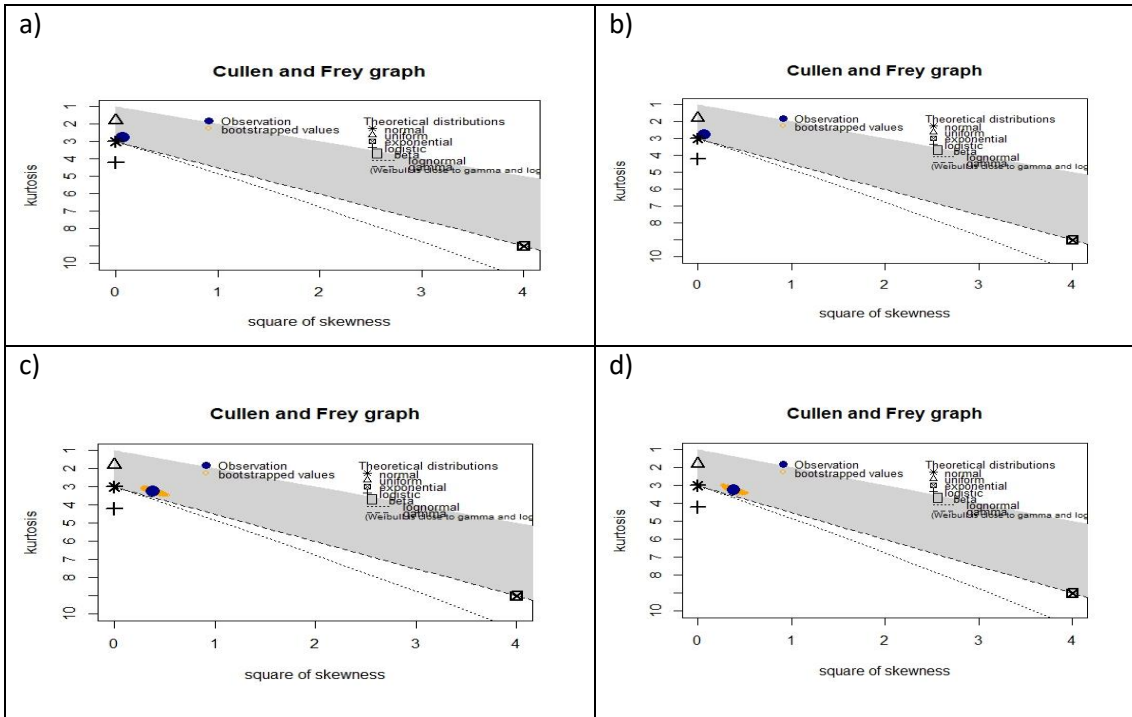
Baroni-Urbani & Buser\_2



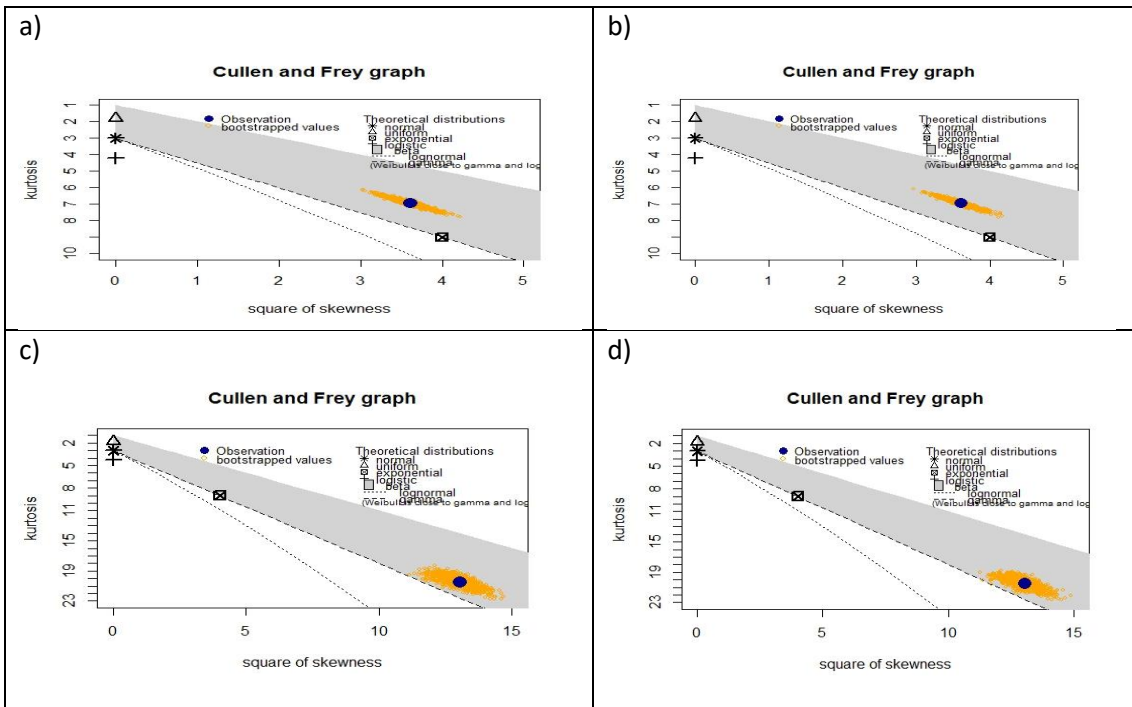
Braun-Blanquet



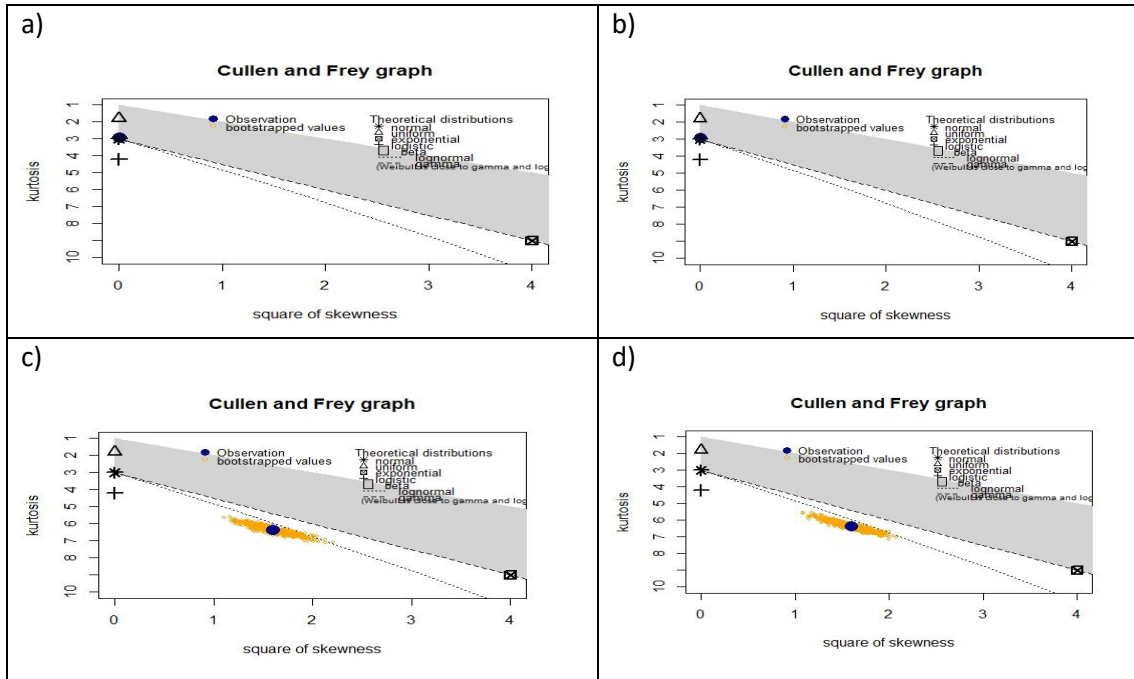
### Clement



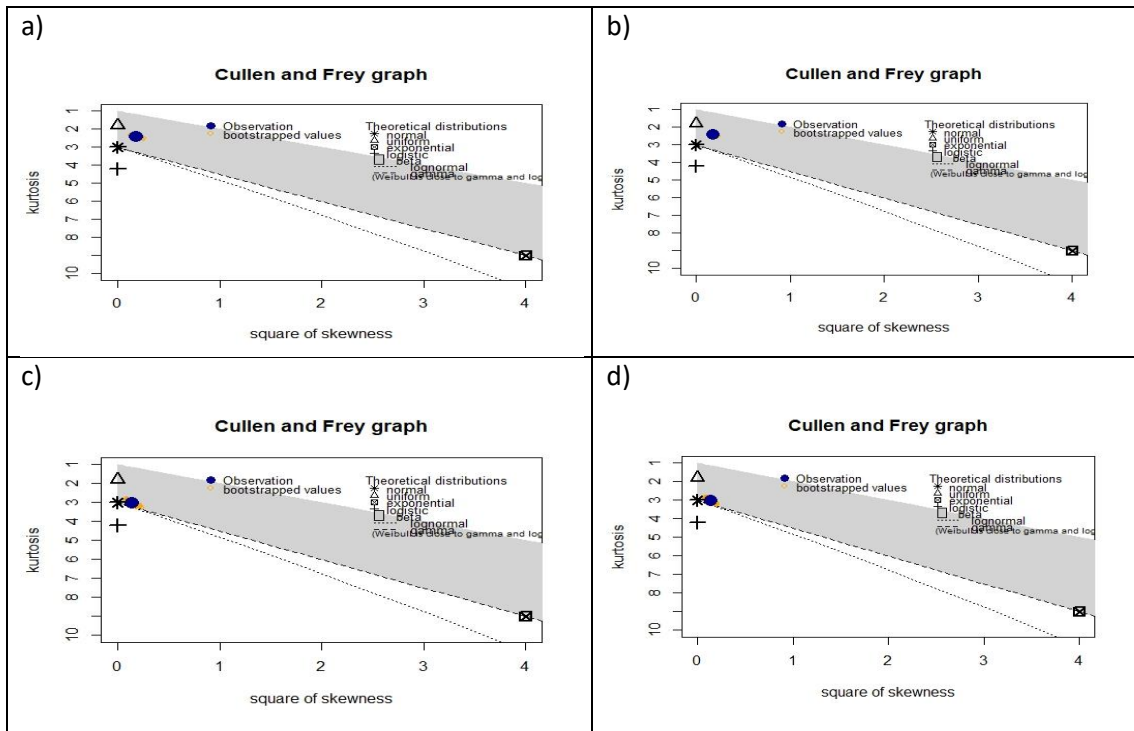
### Clifford & Stephenson



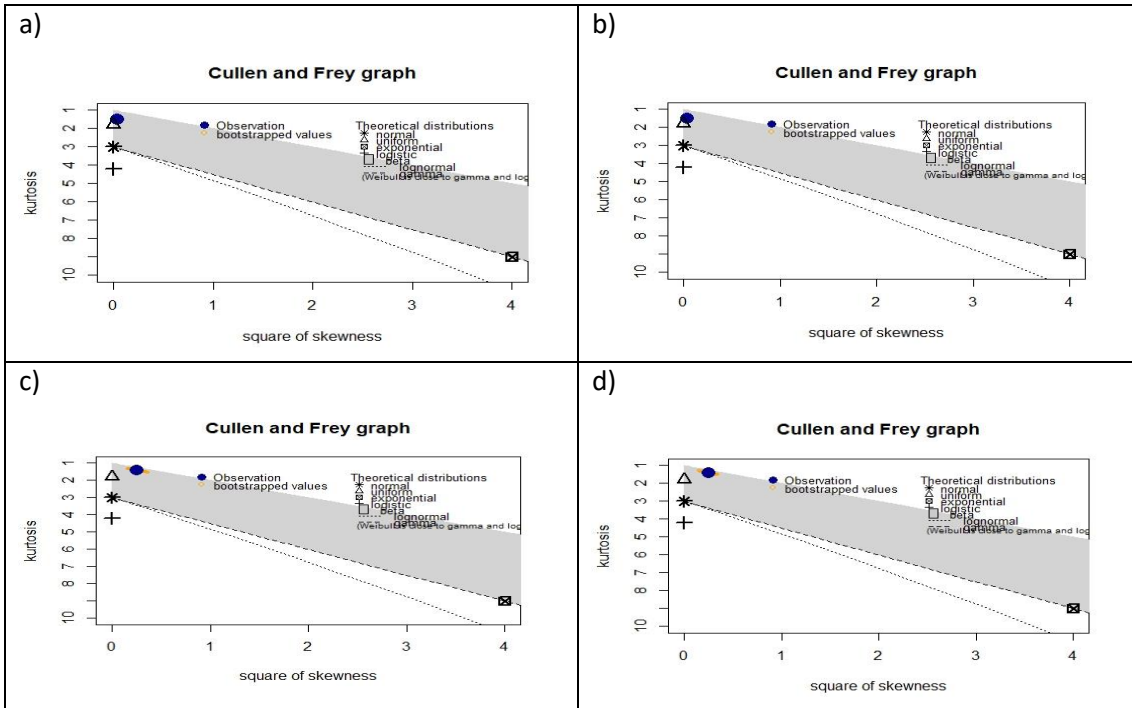
Cohen



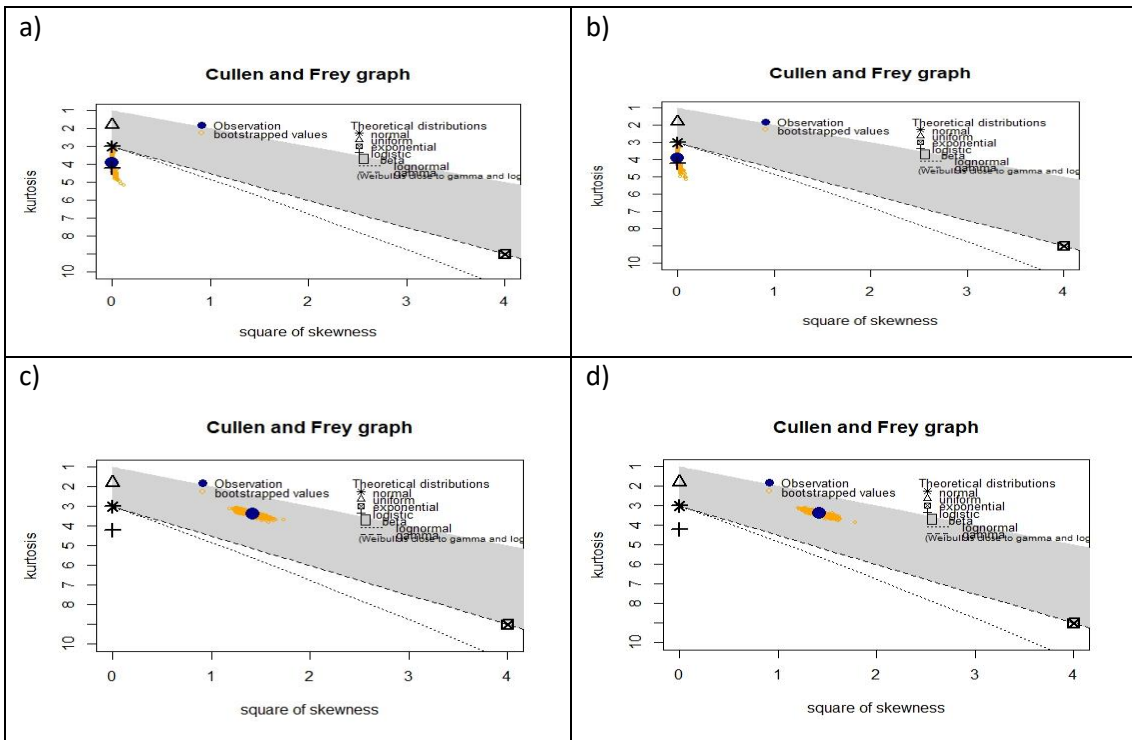
Cole\_1



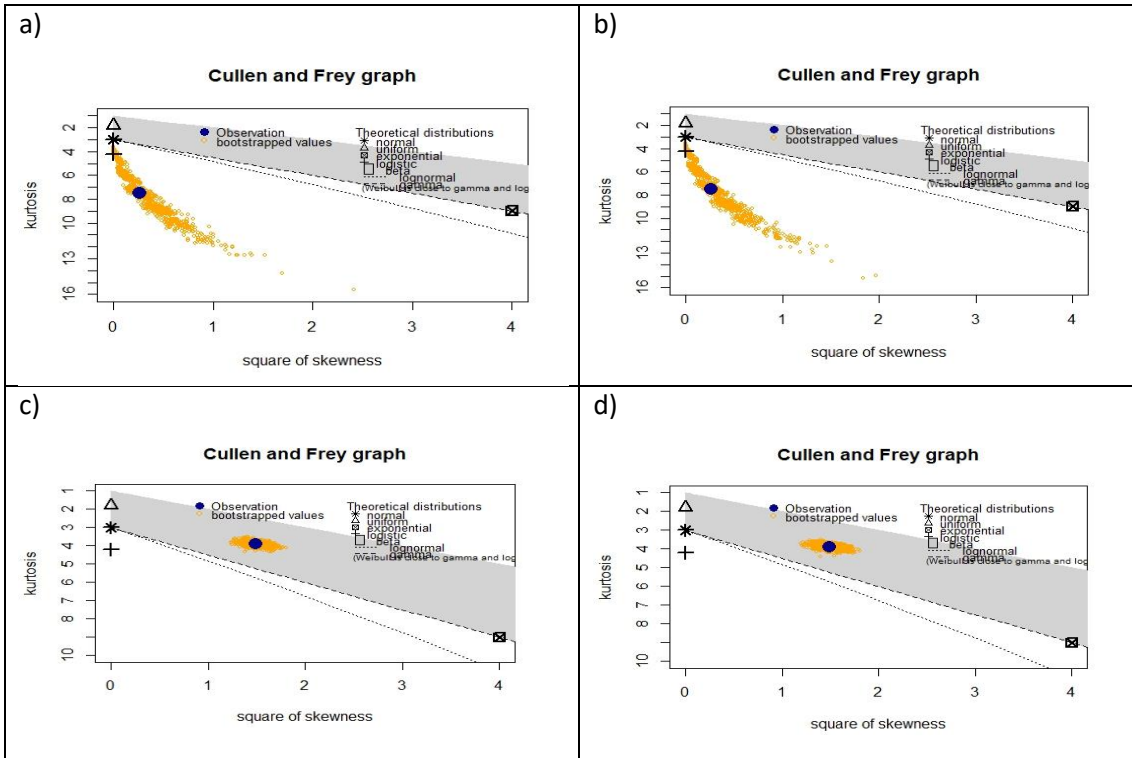
Cole\_2



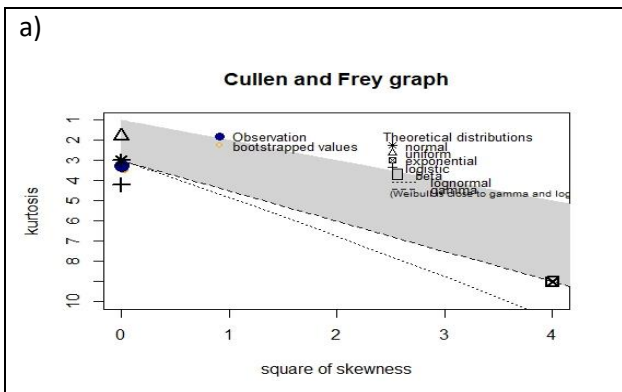
Cole\_3



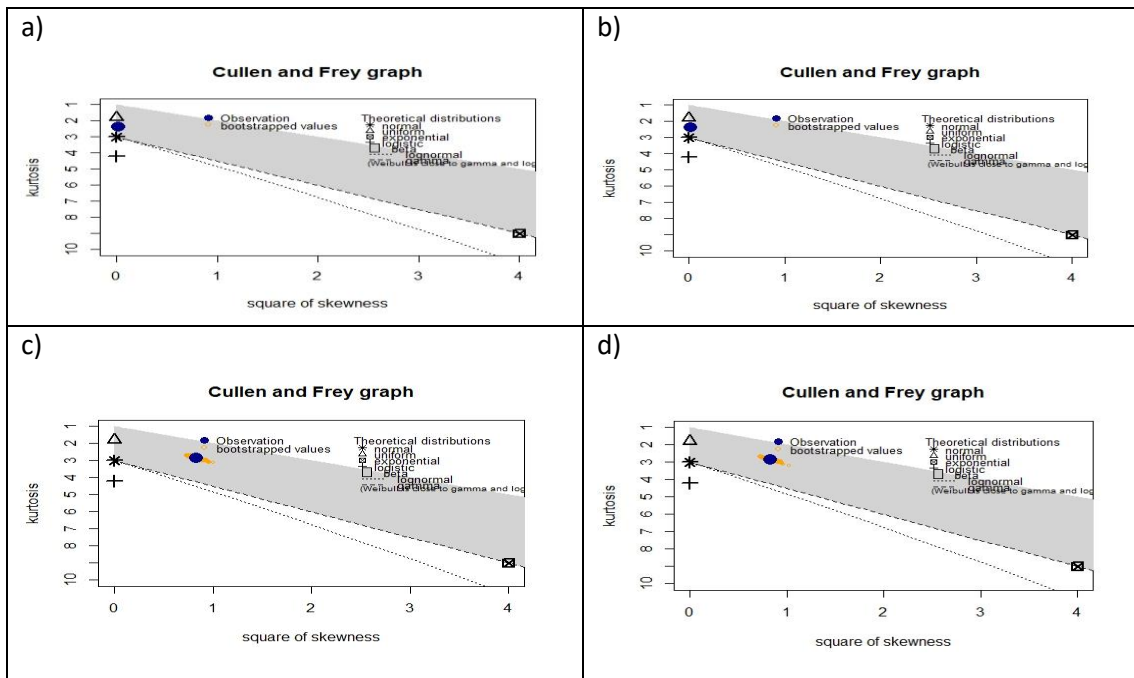
Cole\_4



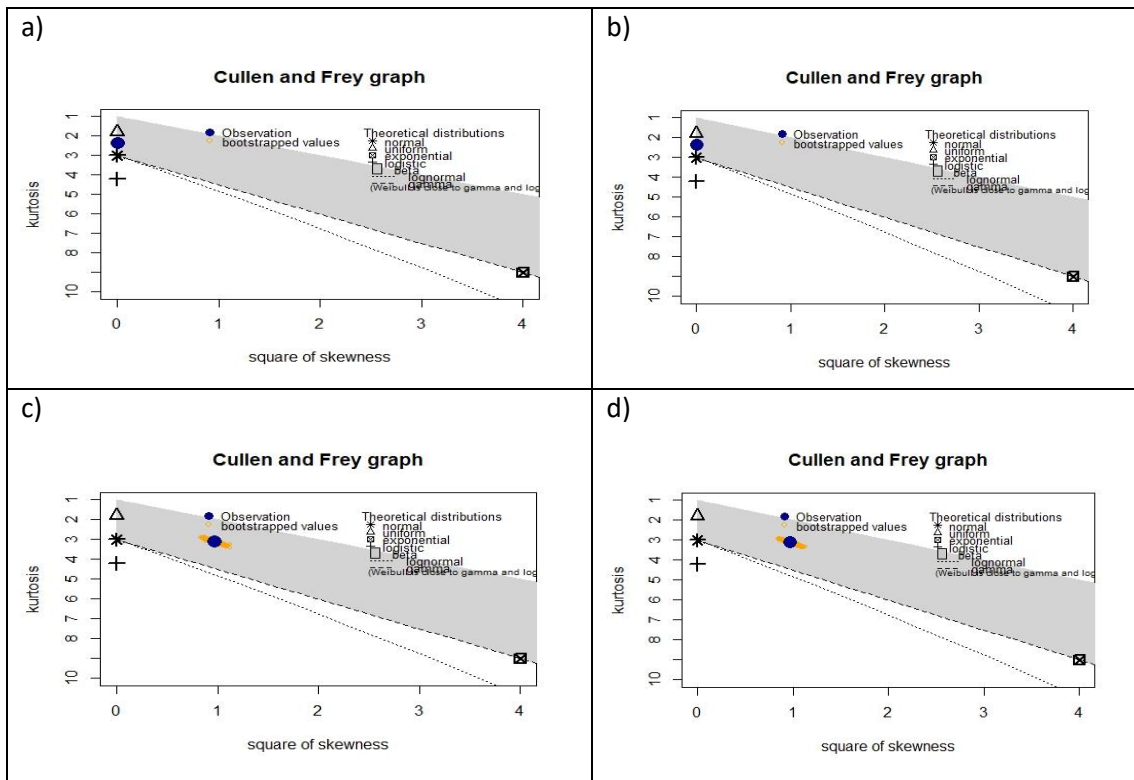
Dennis



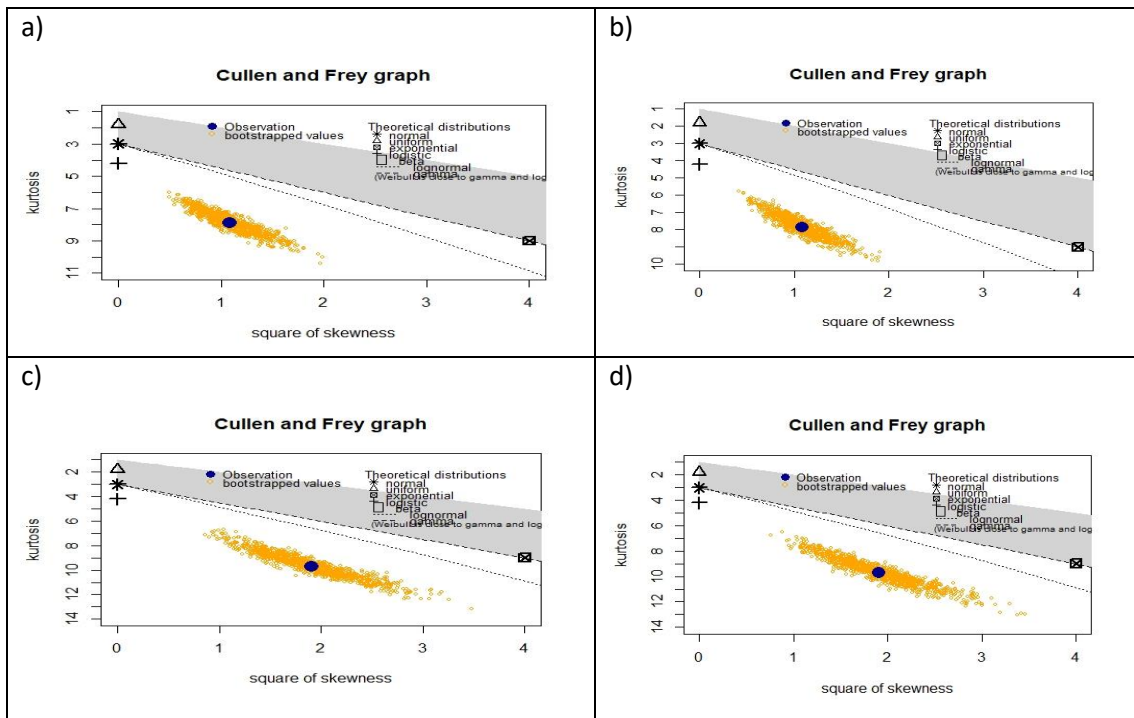
### Dice\_1



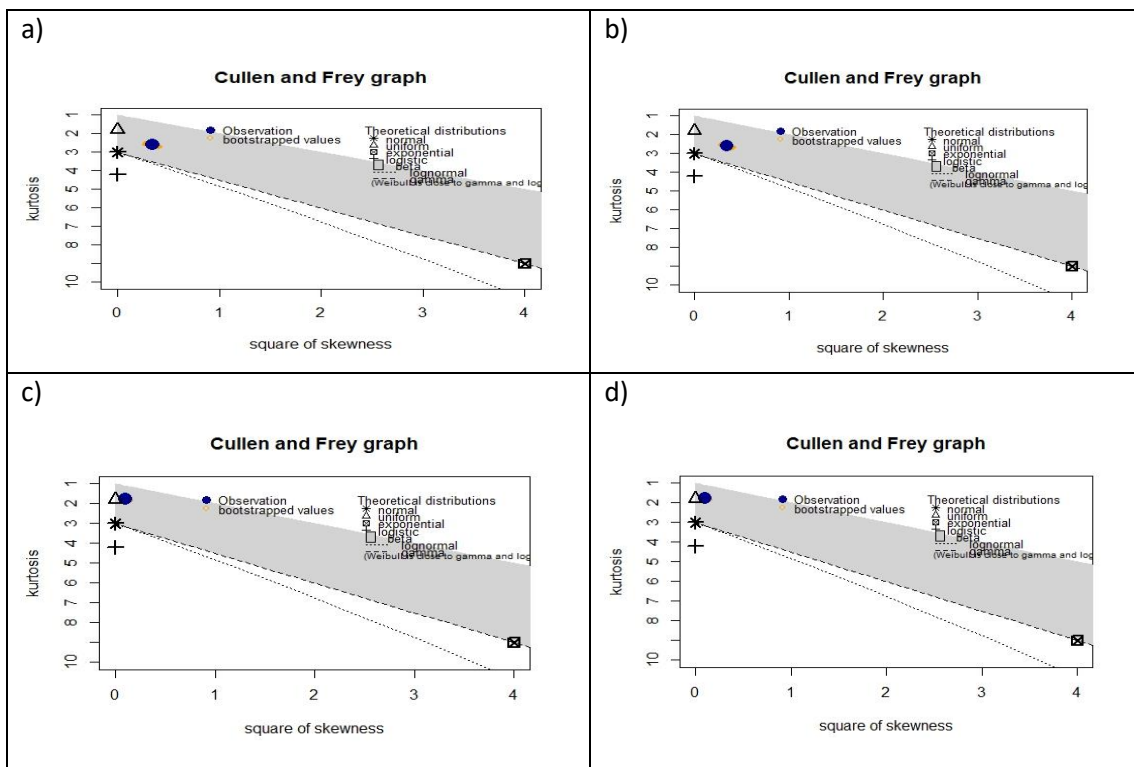
### Dice\_2



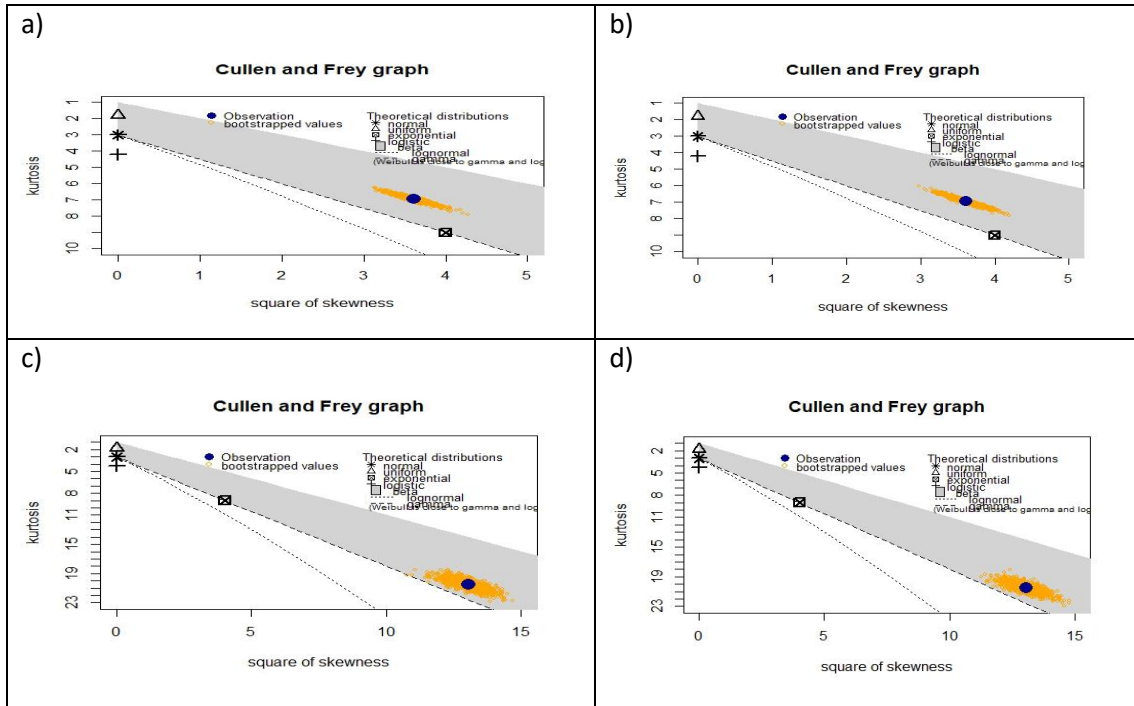
### Dice\_3



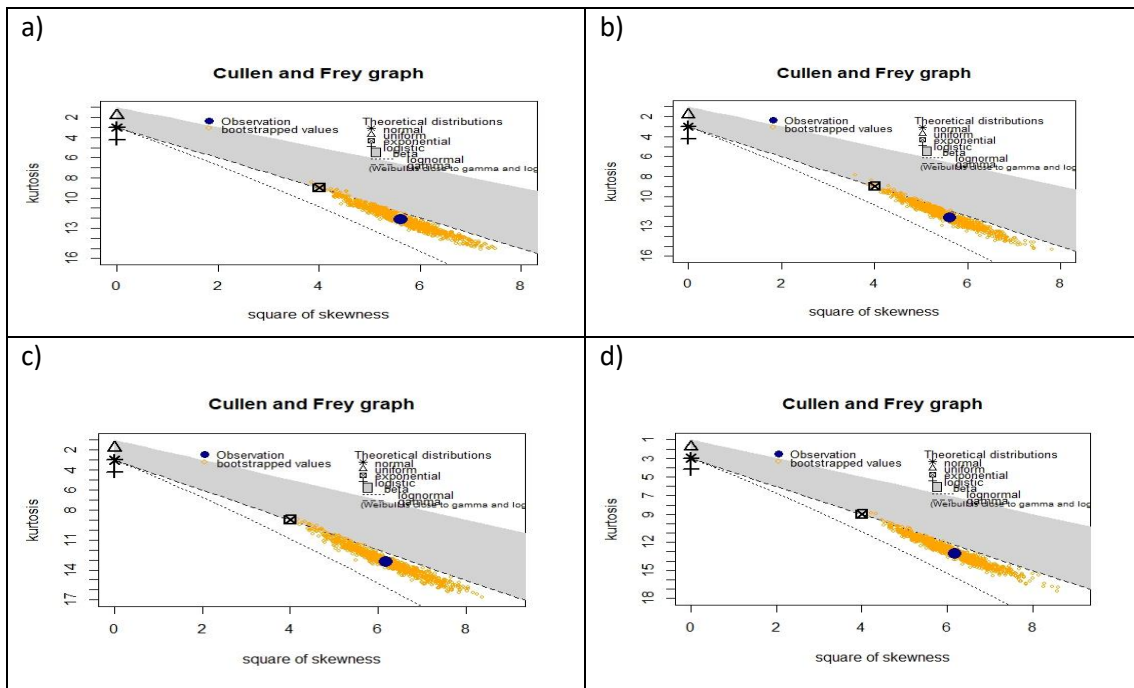
### Digby



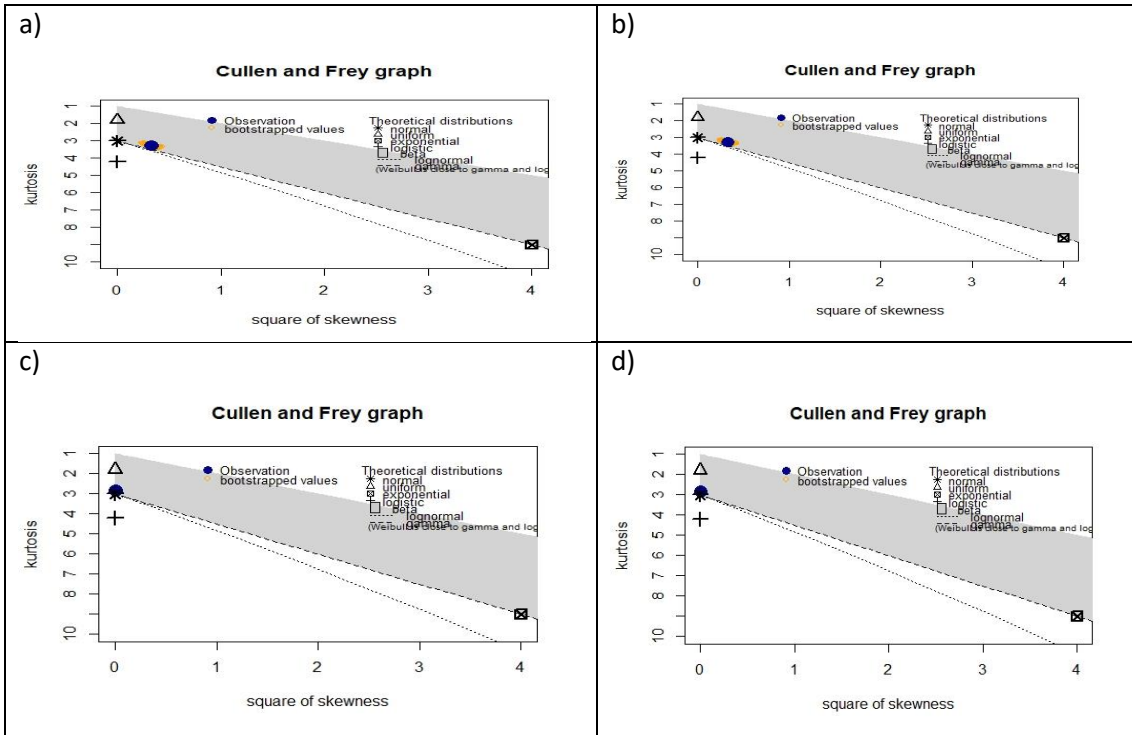
## Doolittle



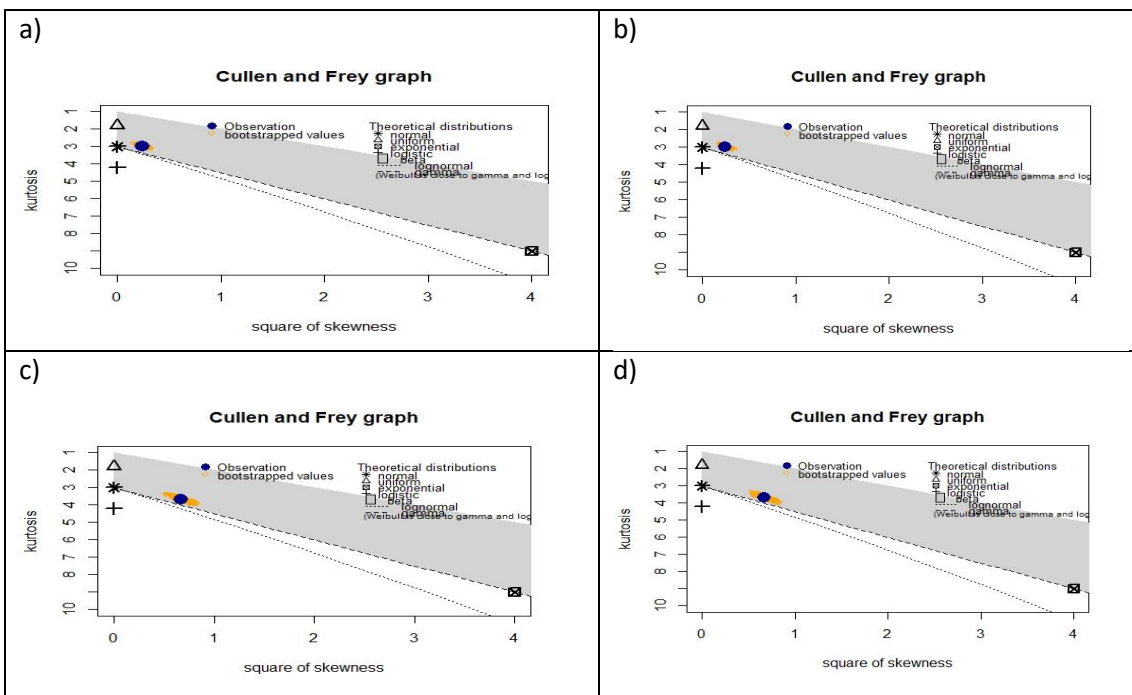
## Eyraud



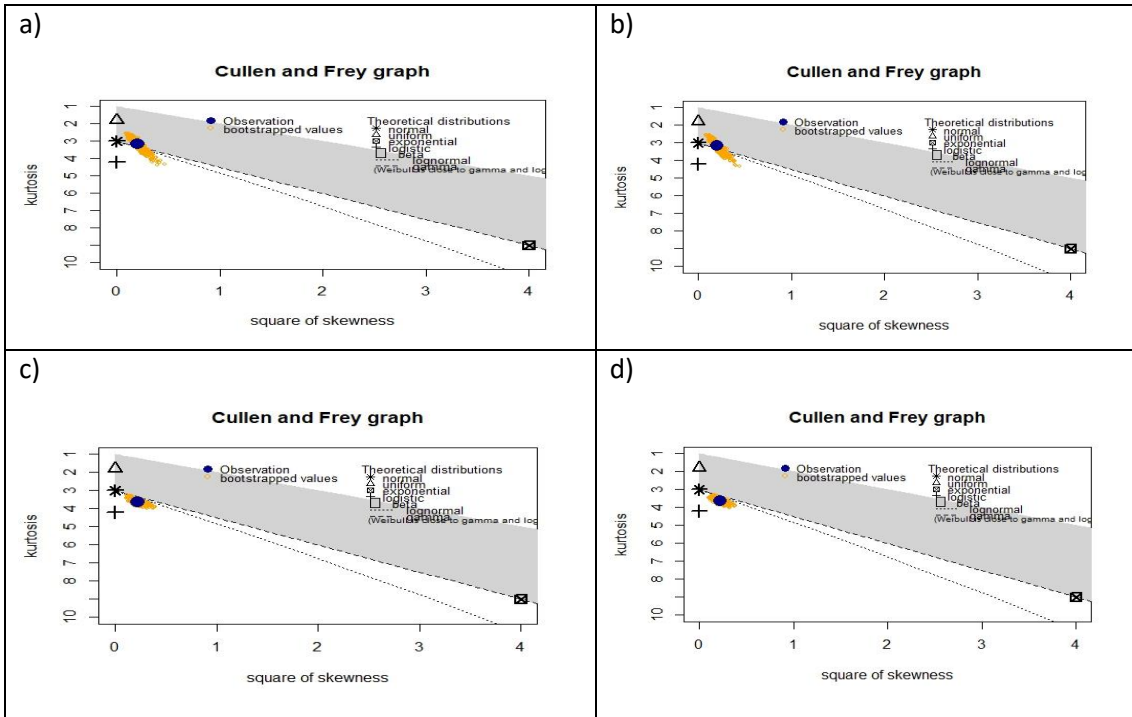
Fager & Mc Gowan



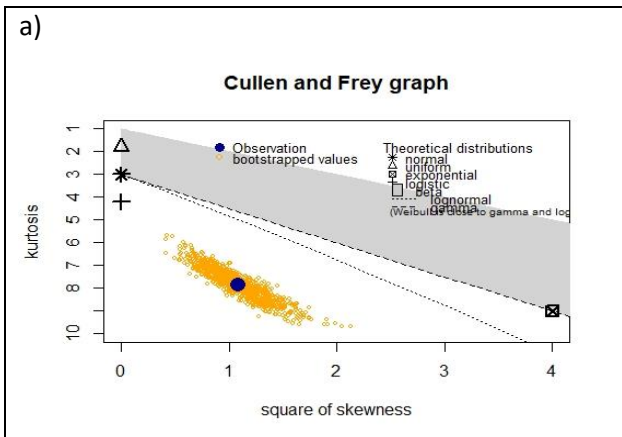
Faith



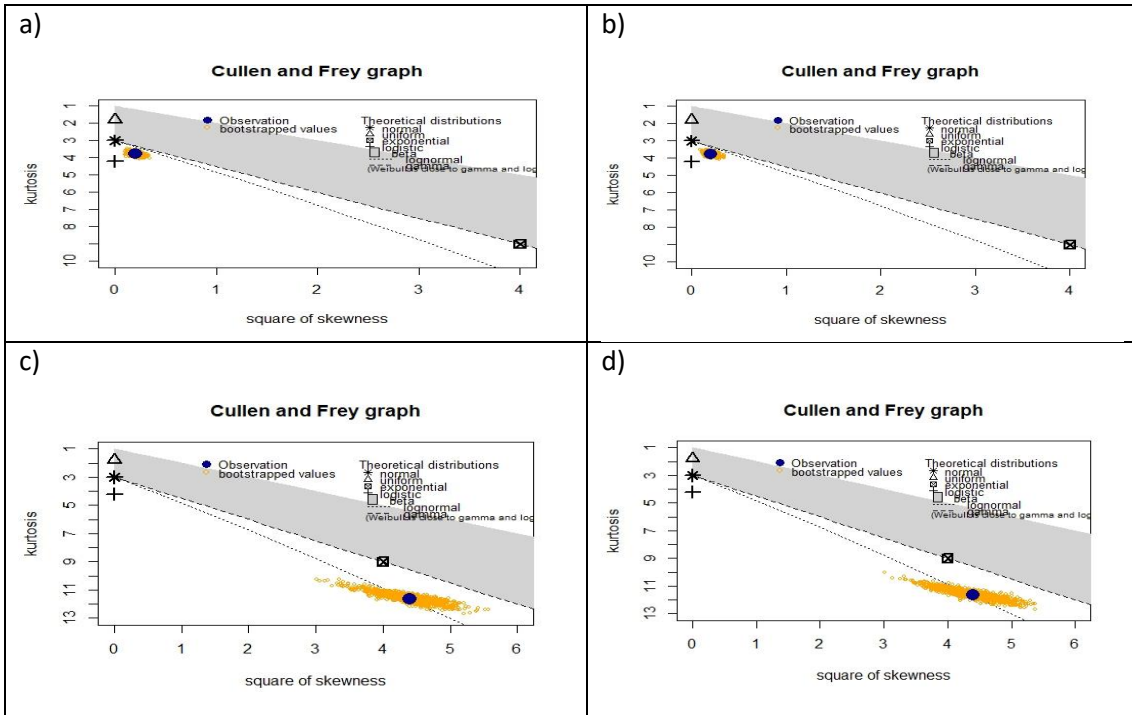
**Fleiss**



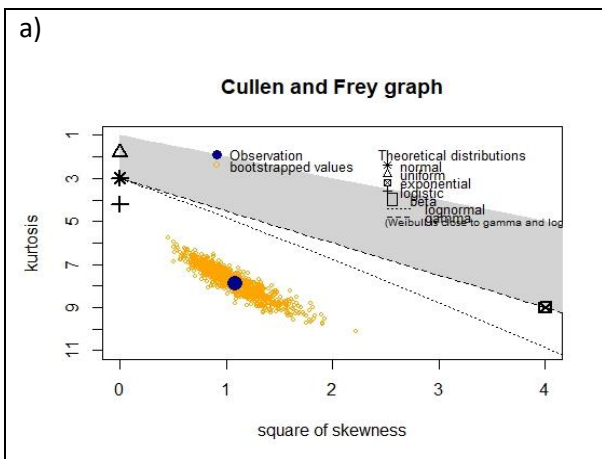
**Forbes\_1**



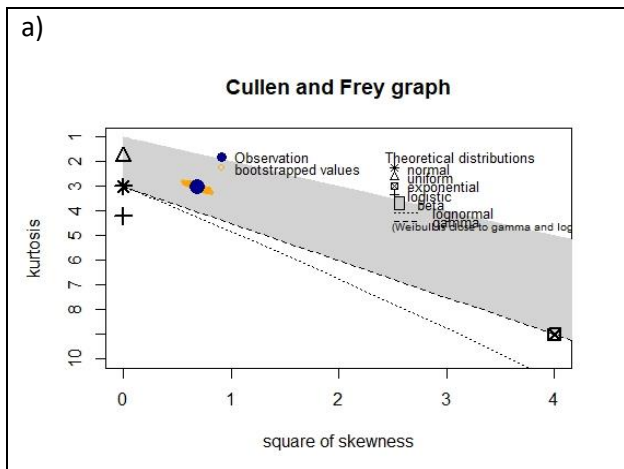
Forbes\_2



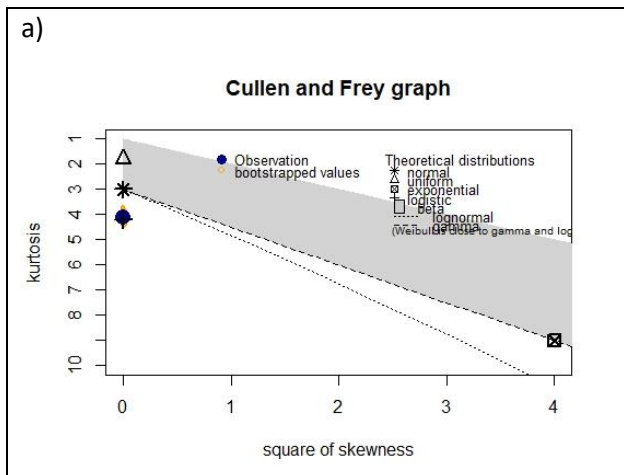
Forbes\_3



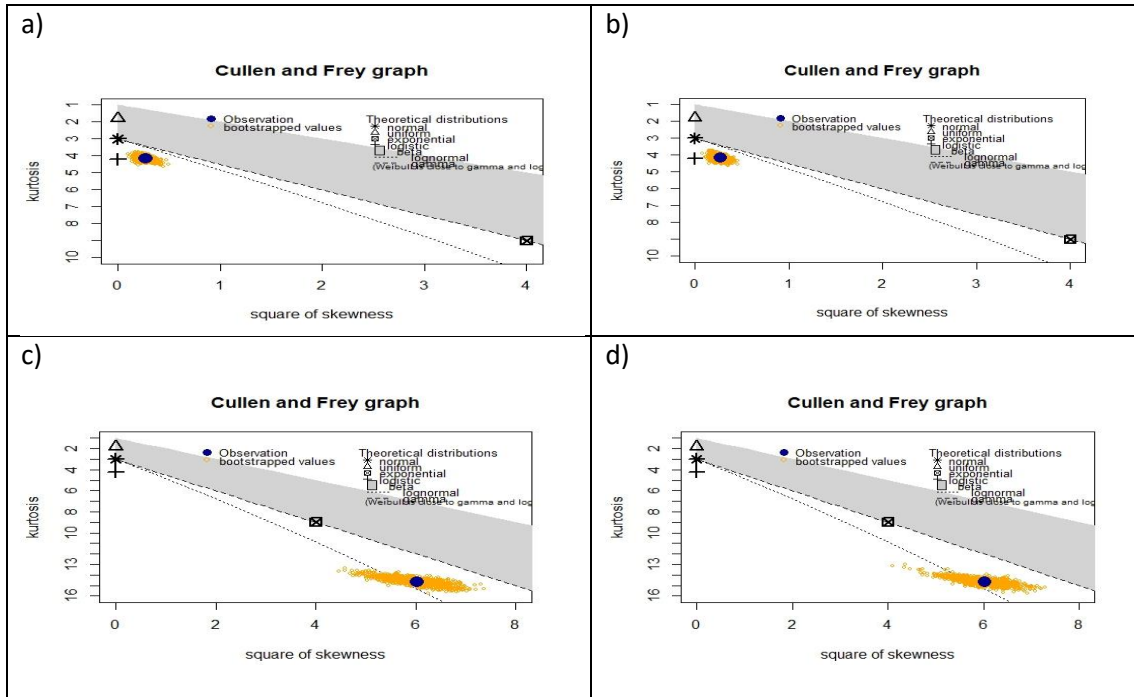
### Fossum



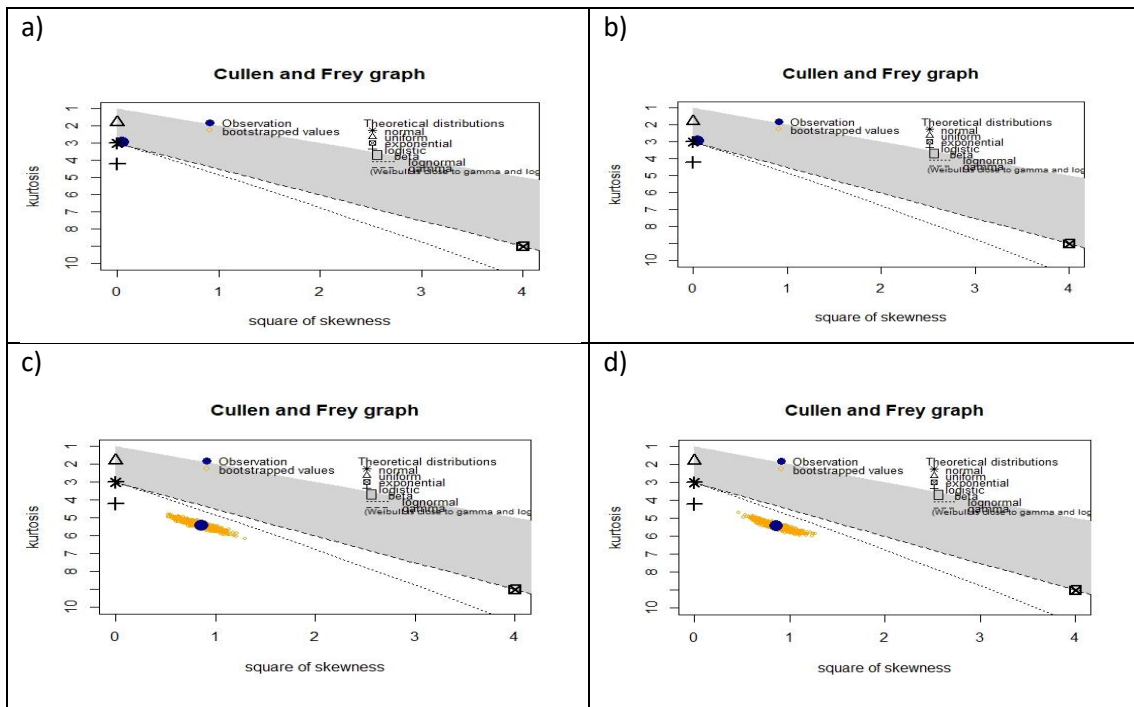
### Gilbert & Wells



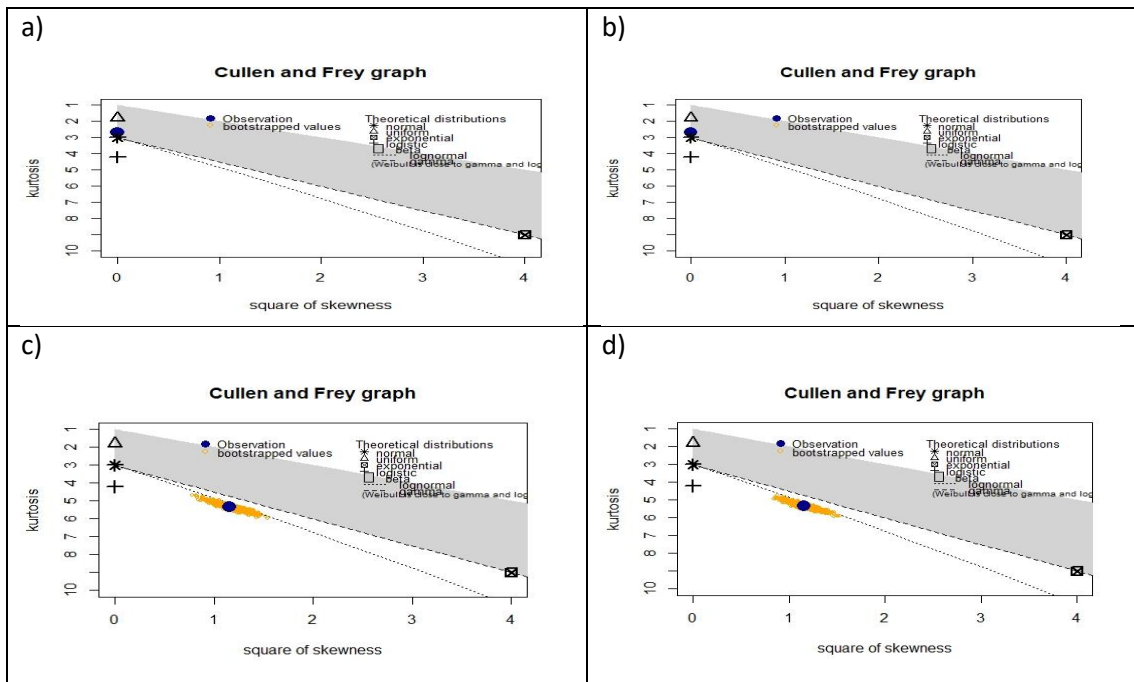
## Goodall



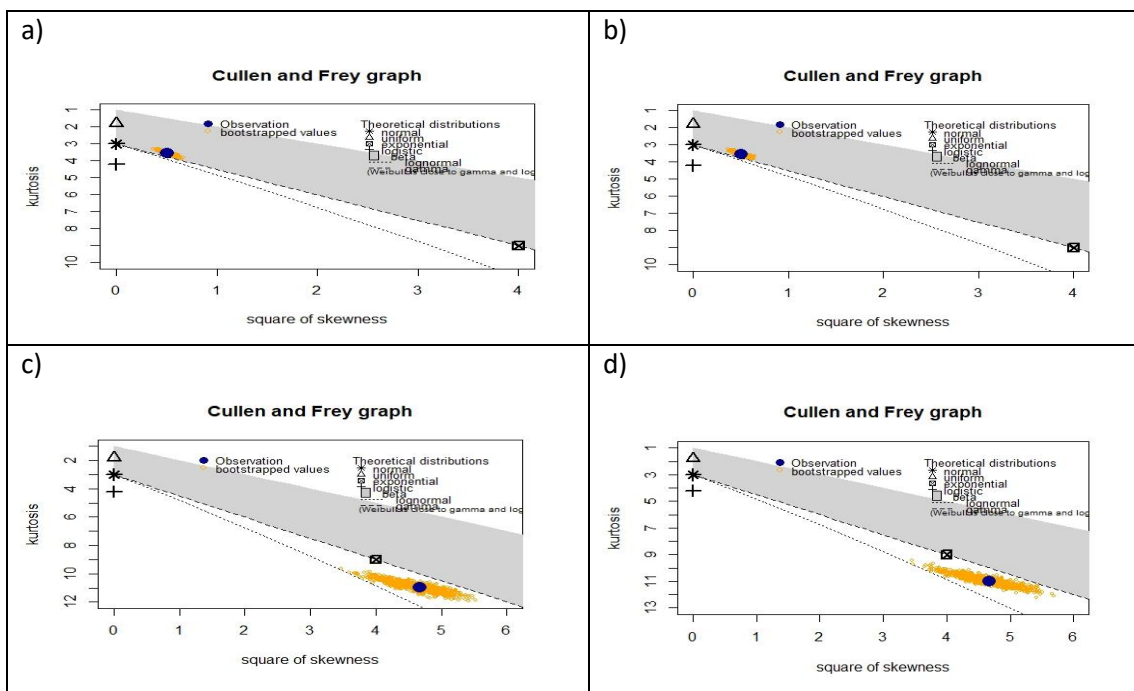
## Goodman & Kruskal



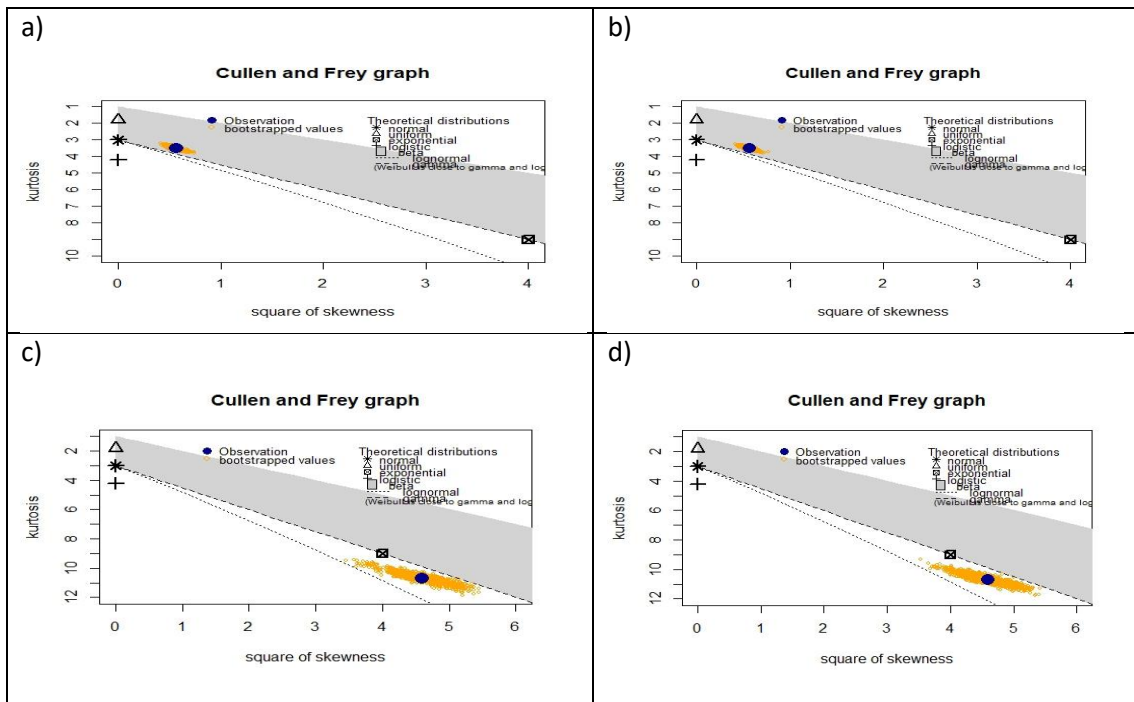
### Hamann



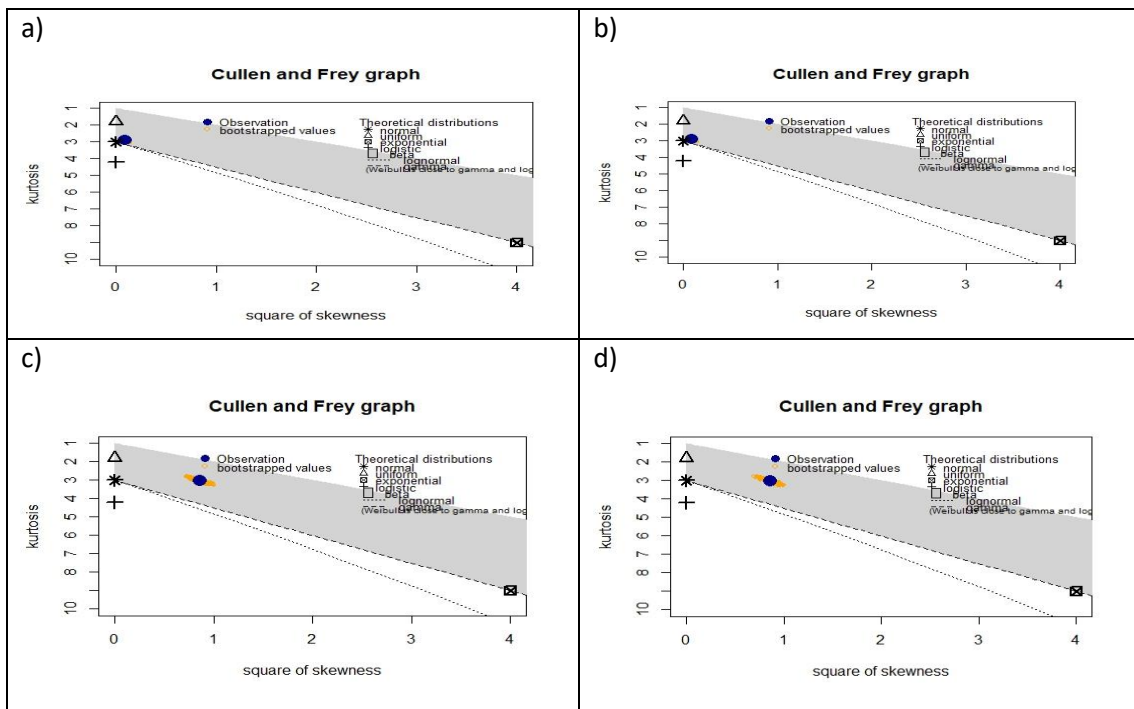
### Harris & Lahey



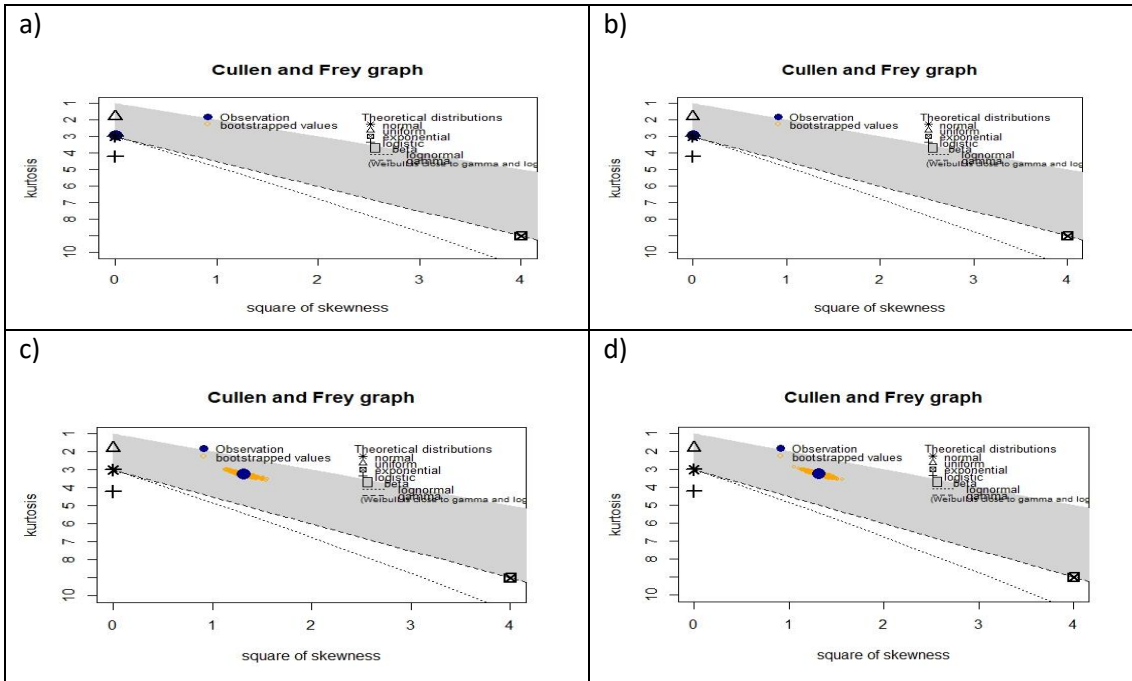
## Hawkins & Dotson



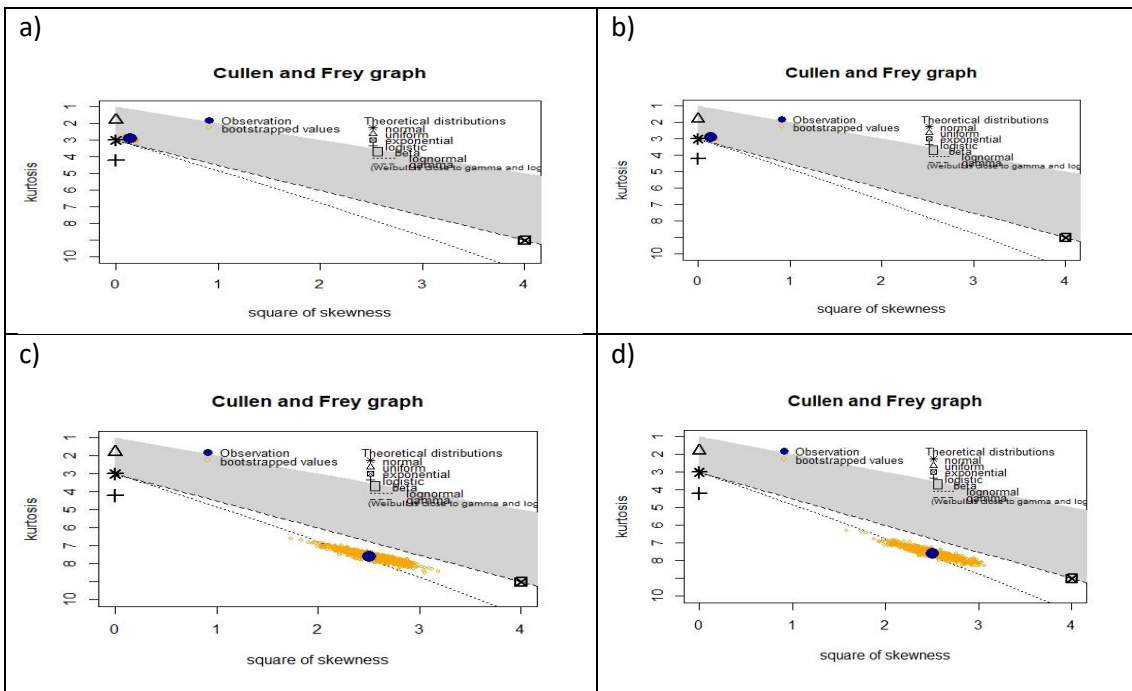
## Hurlbert\_1



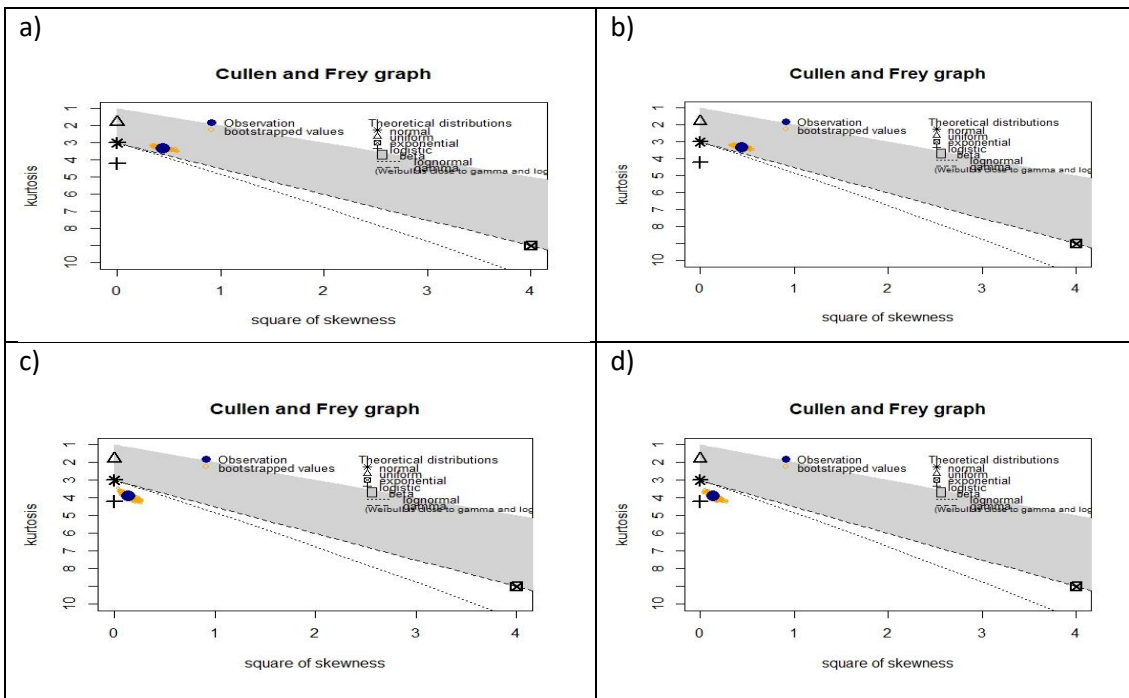
## Hurlbert\_2



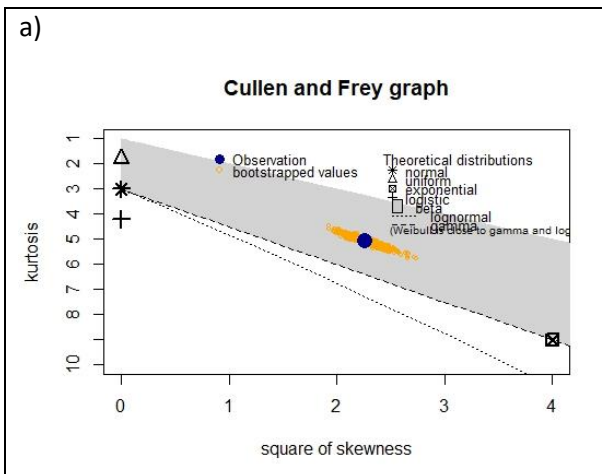
## Jaccard



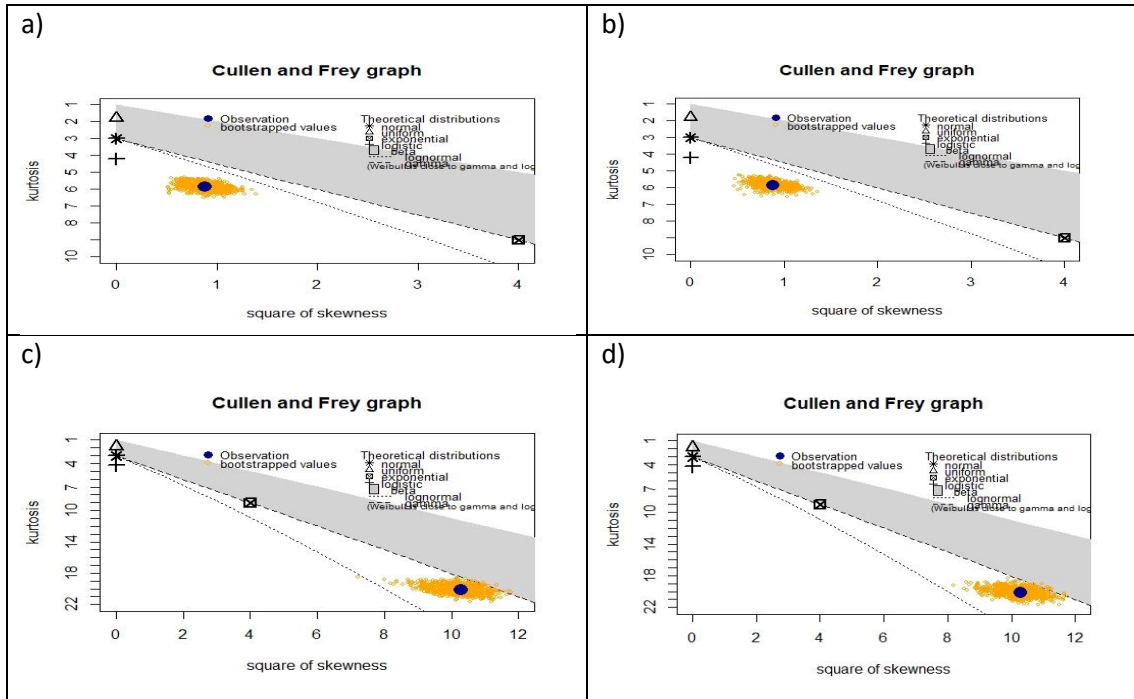
Johnson\_1



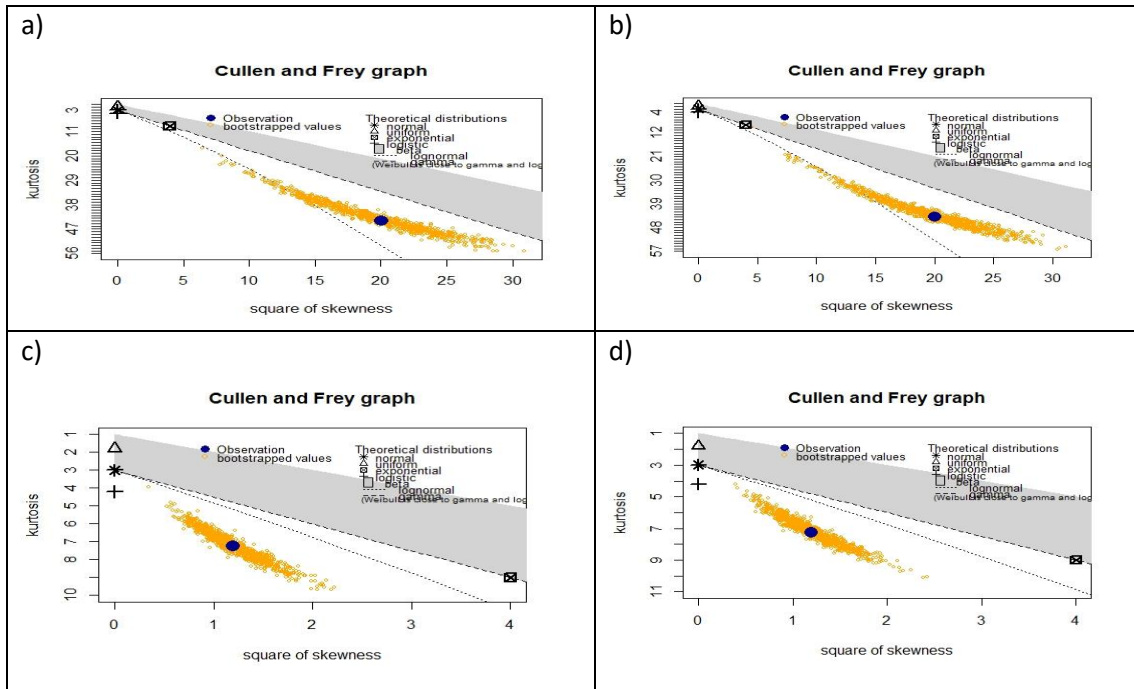
Johnson\_2



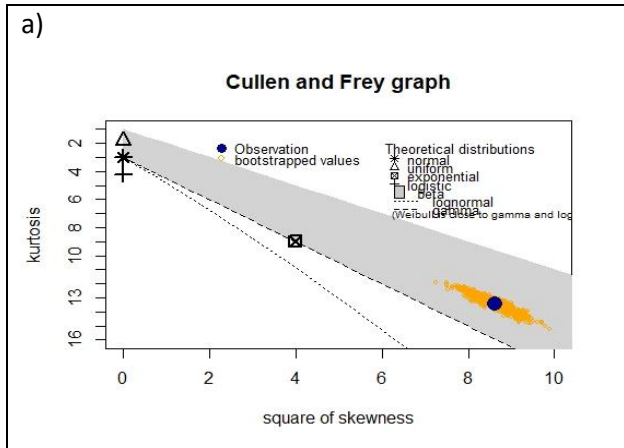
## Kendall



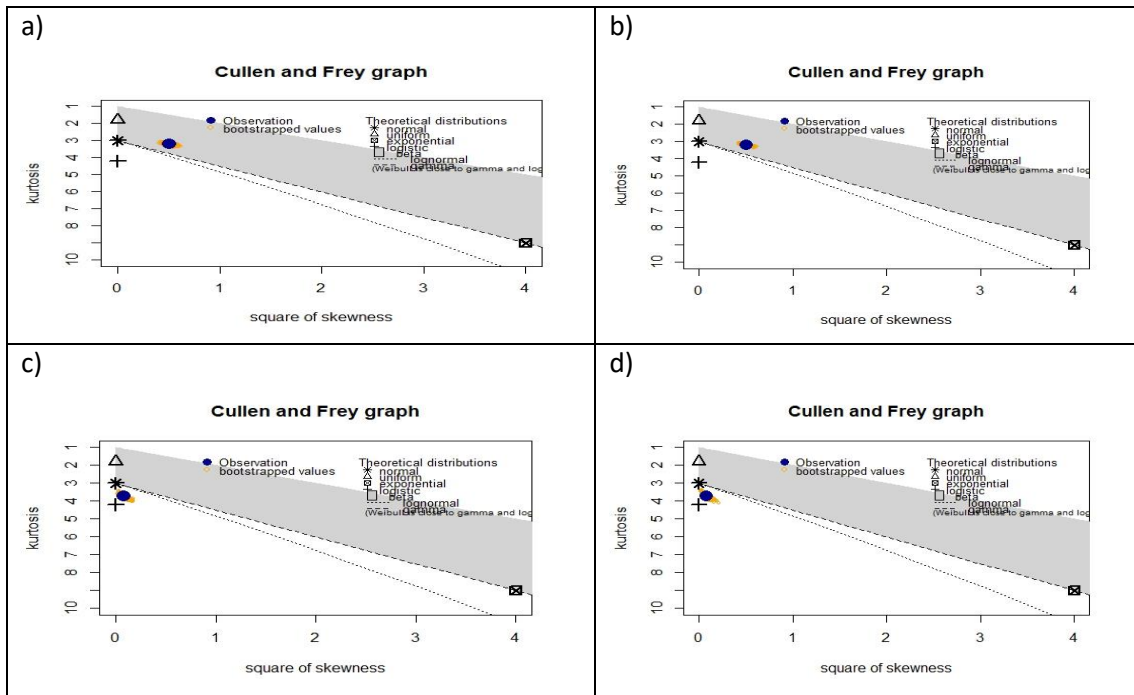
## Kuder & Richardson



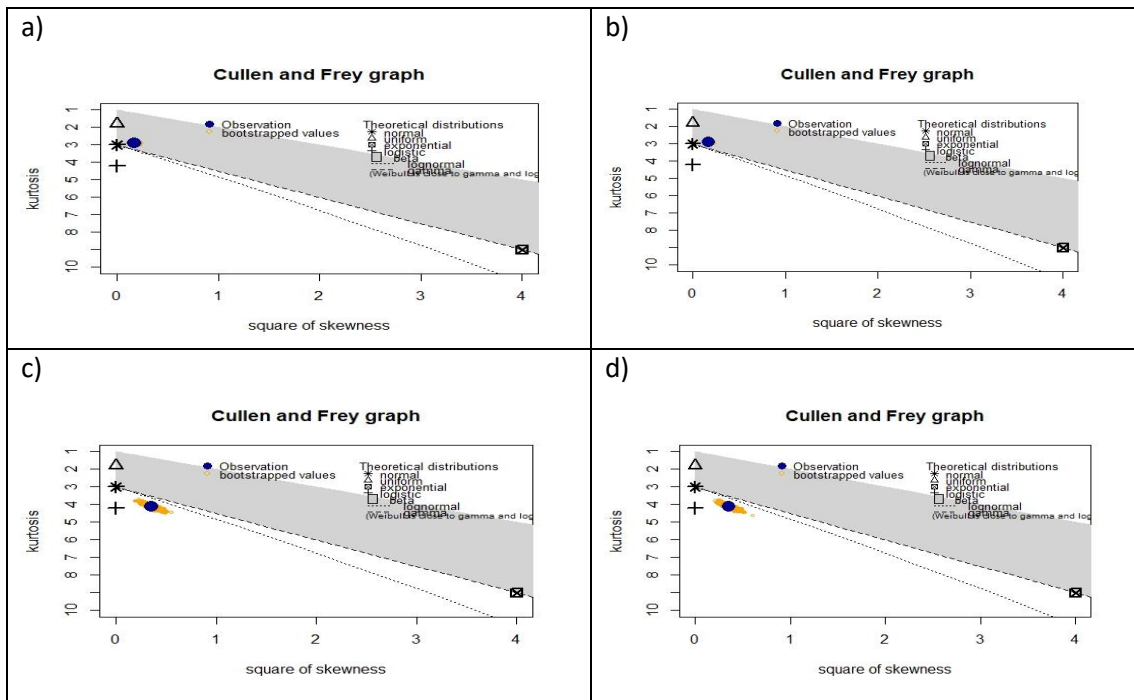
### Kulczynski\_1



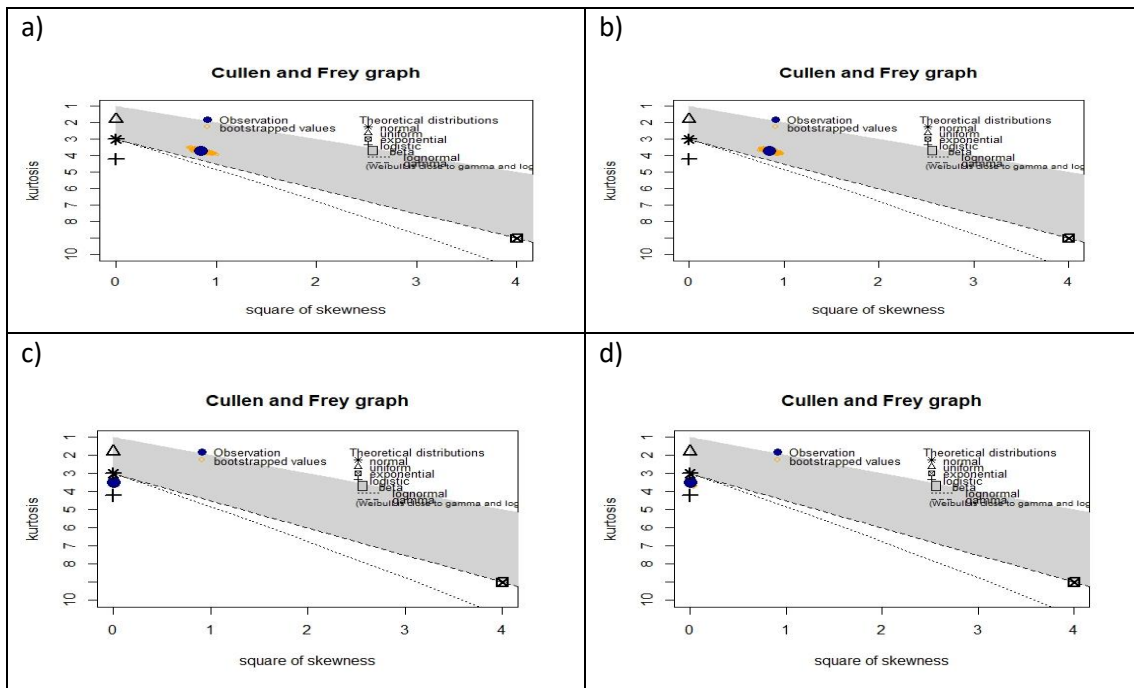
### Kulczynski\_2



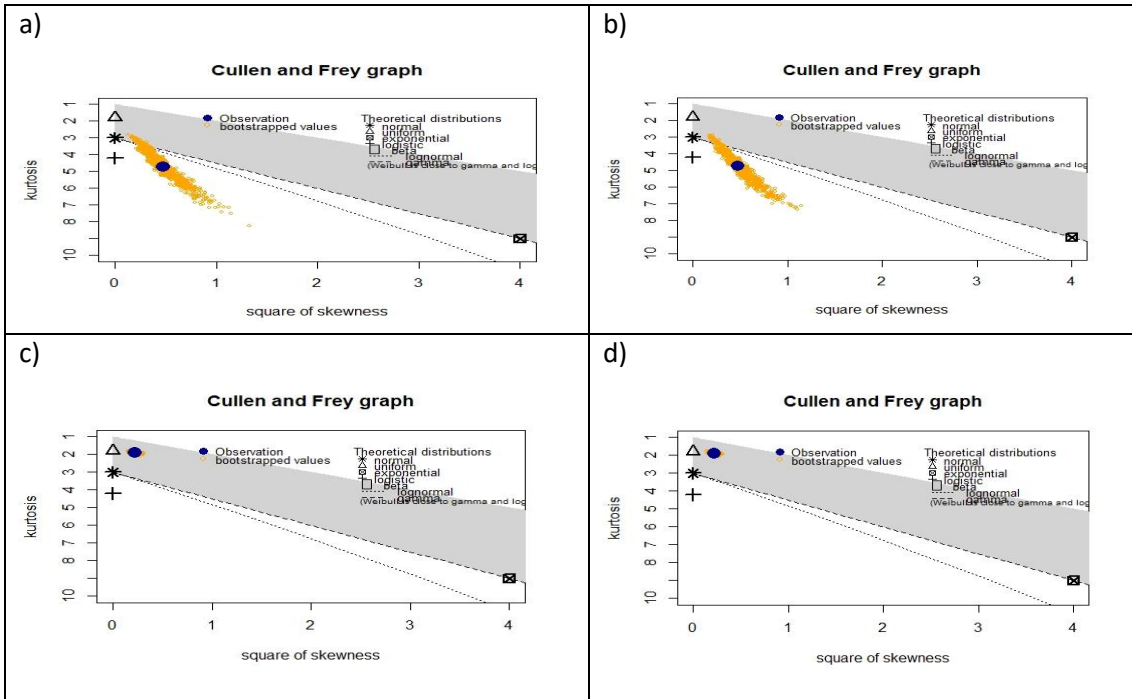
## Lamont & Grant



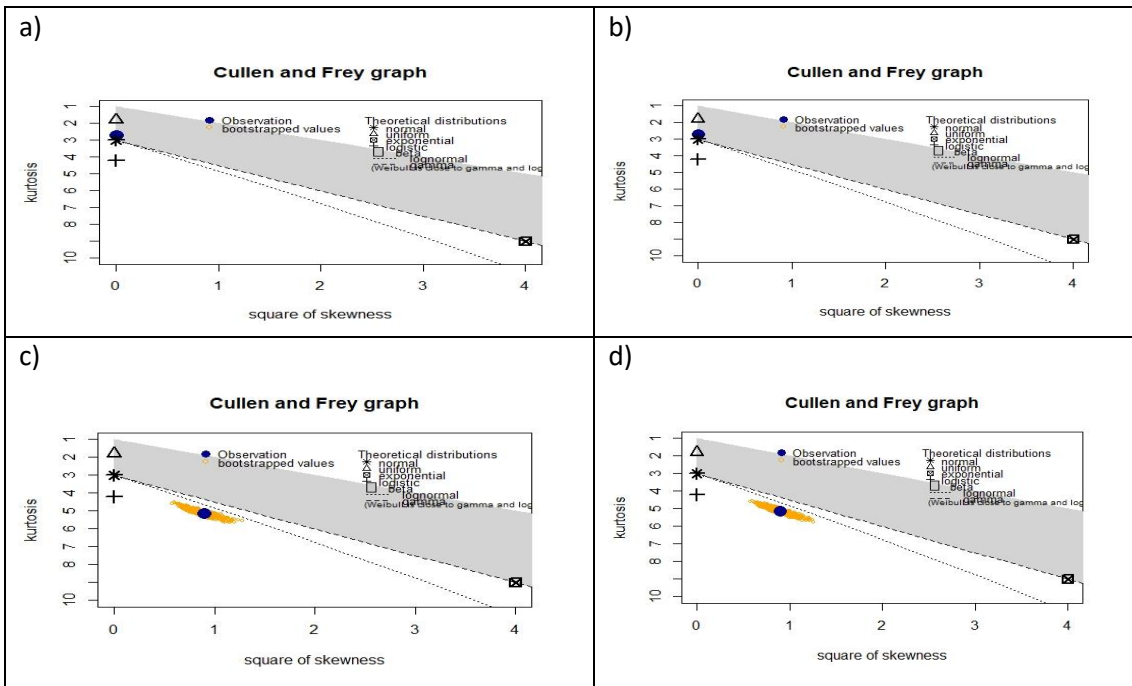
## Legendre & Legendre



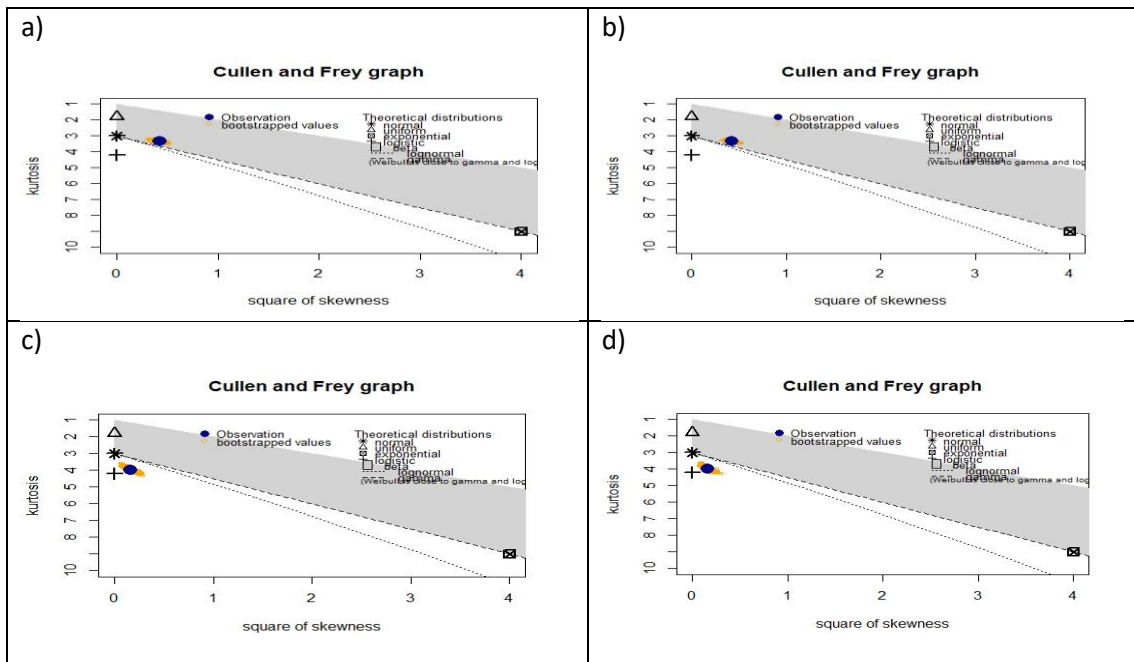
Loevinger



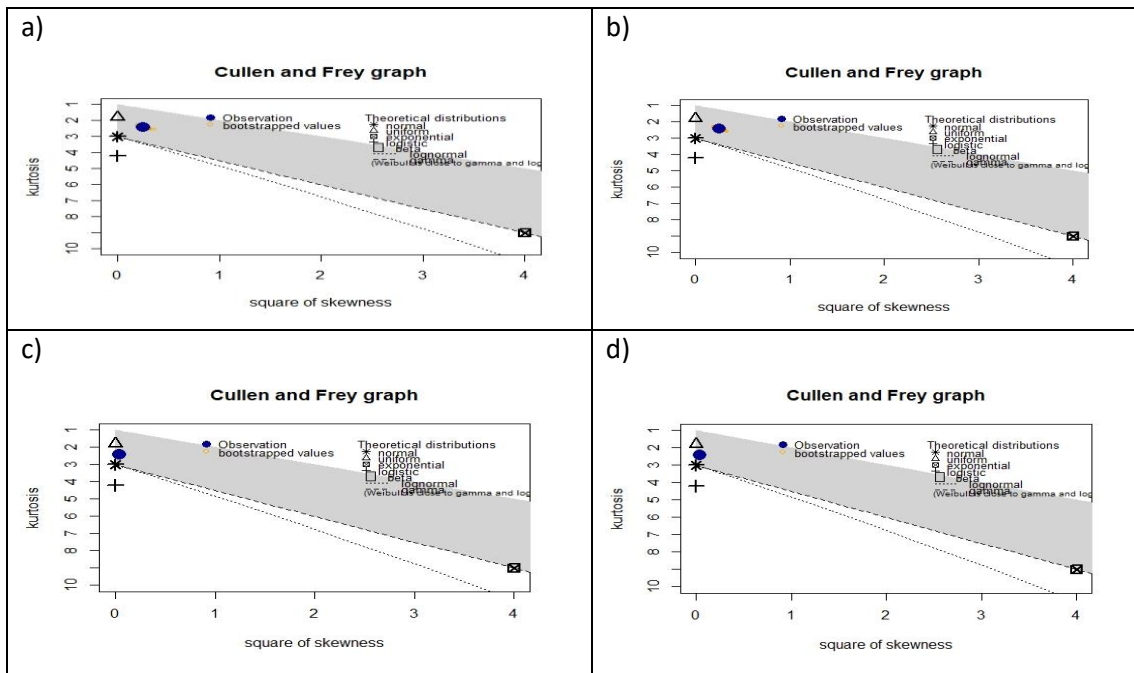
Maxwell & Pilliner



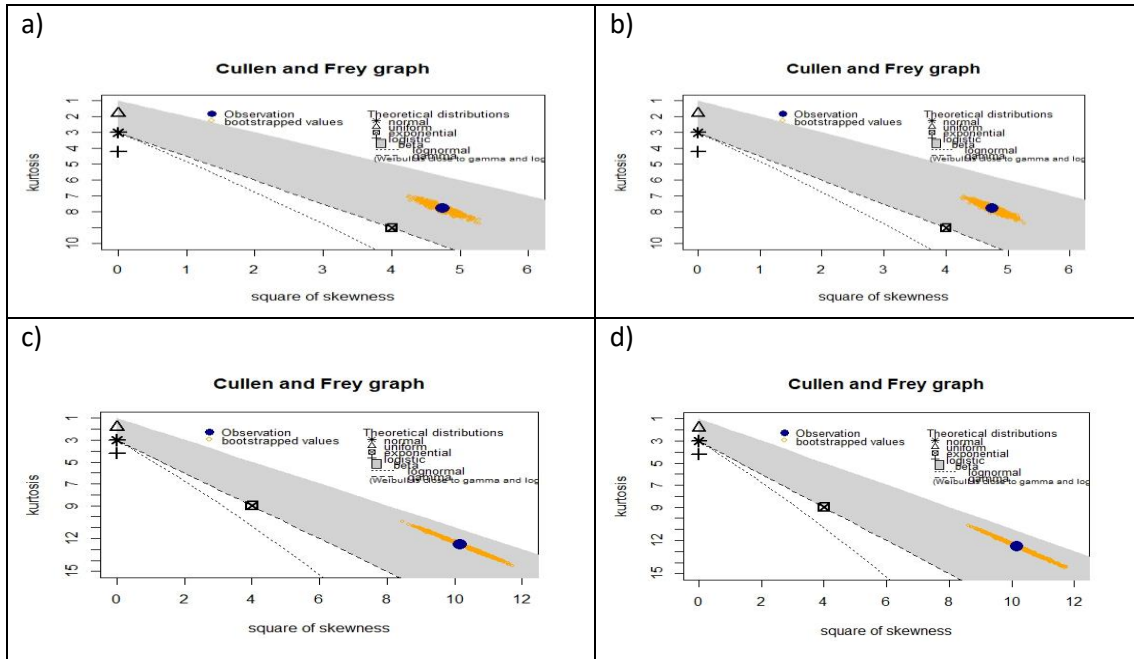
## McConnaughey



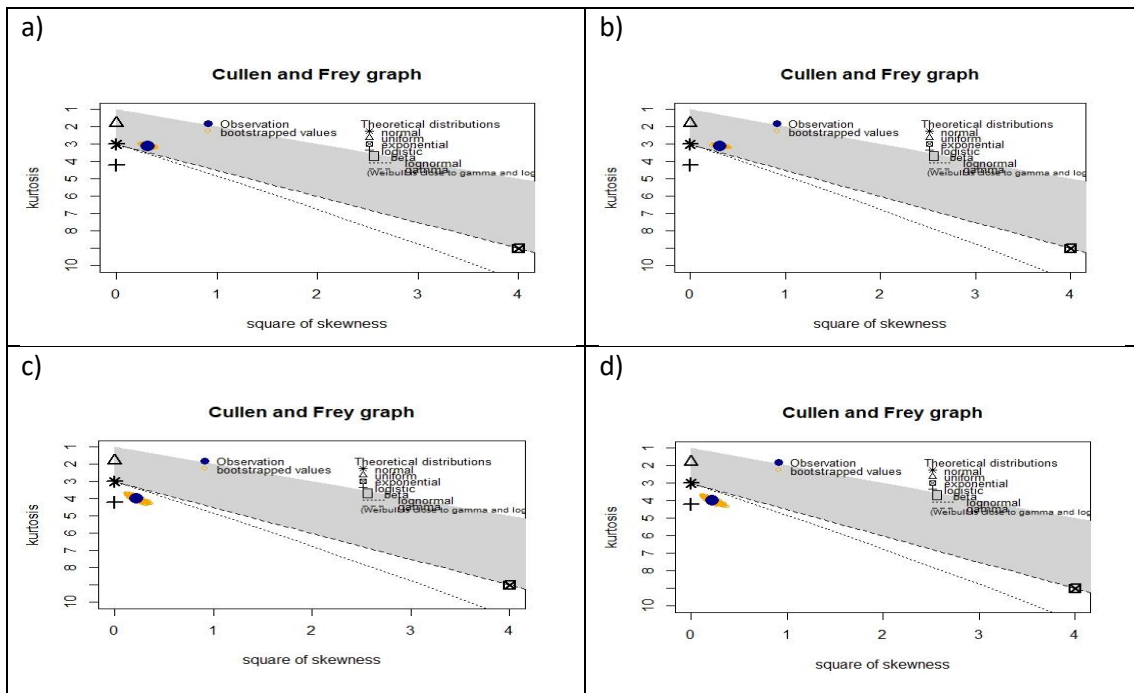
## Michael



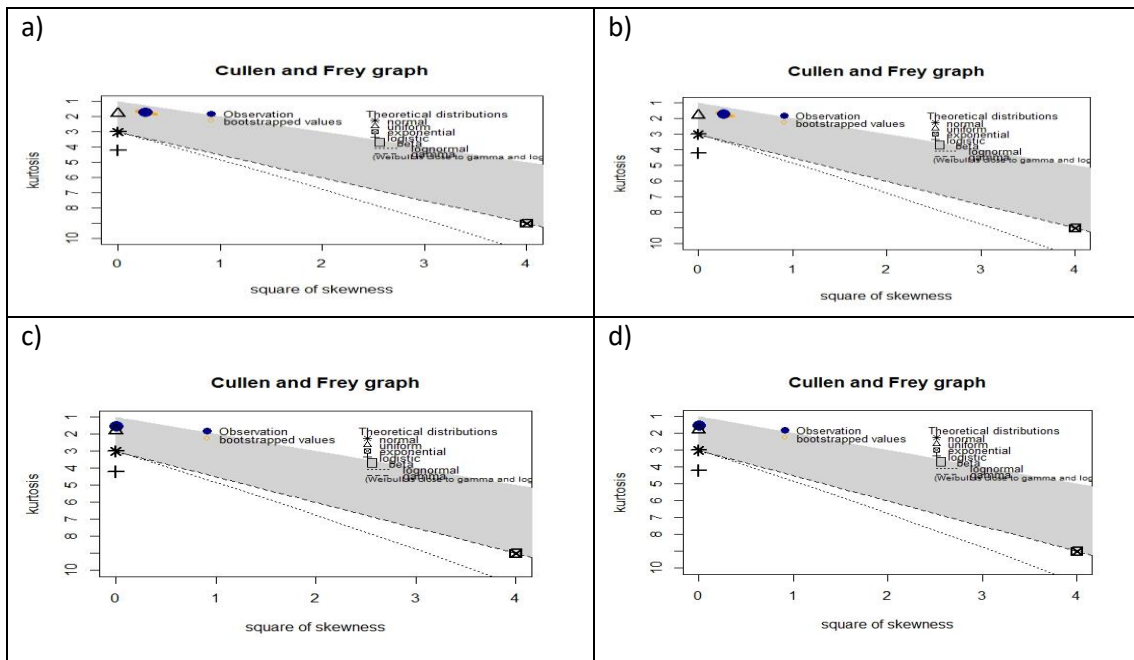
## Mountford



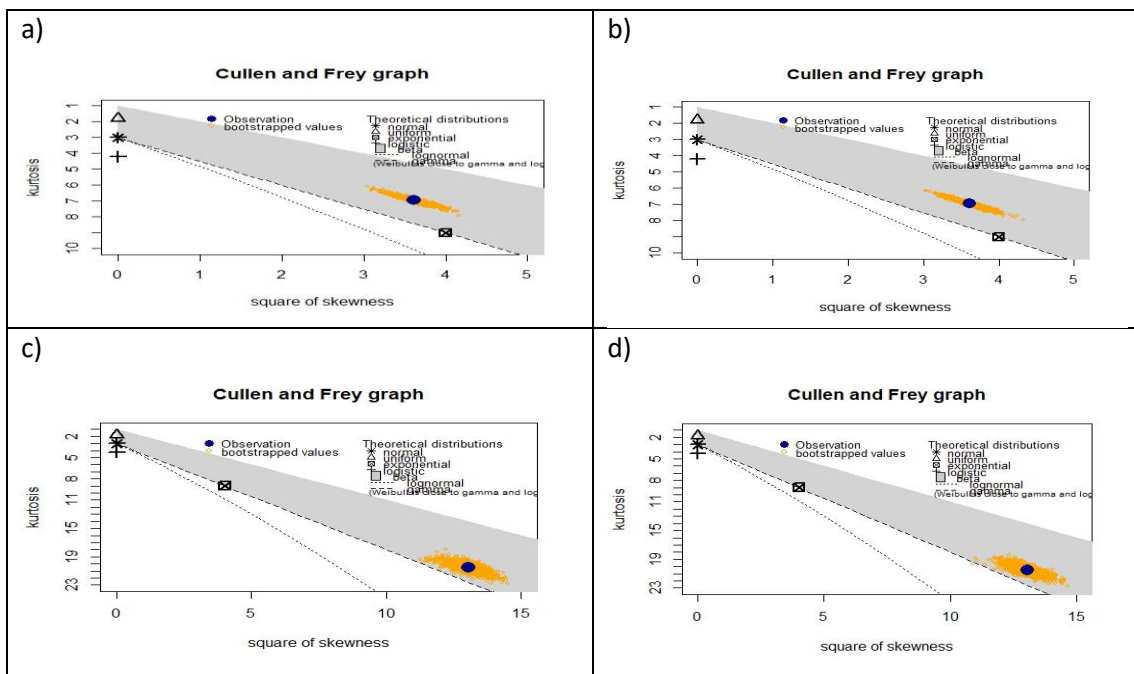
## Ochiai



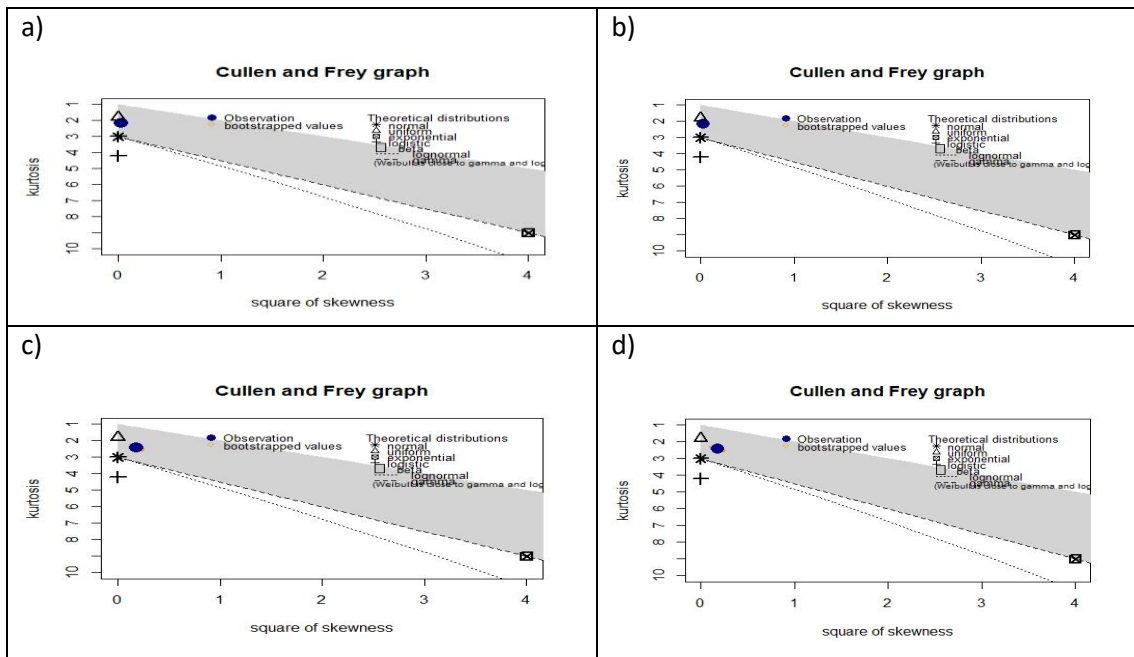
## Pearson & Heron



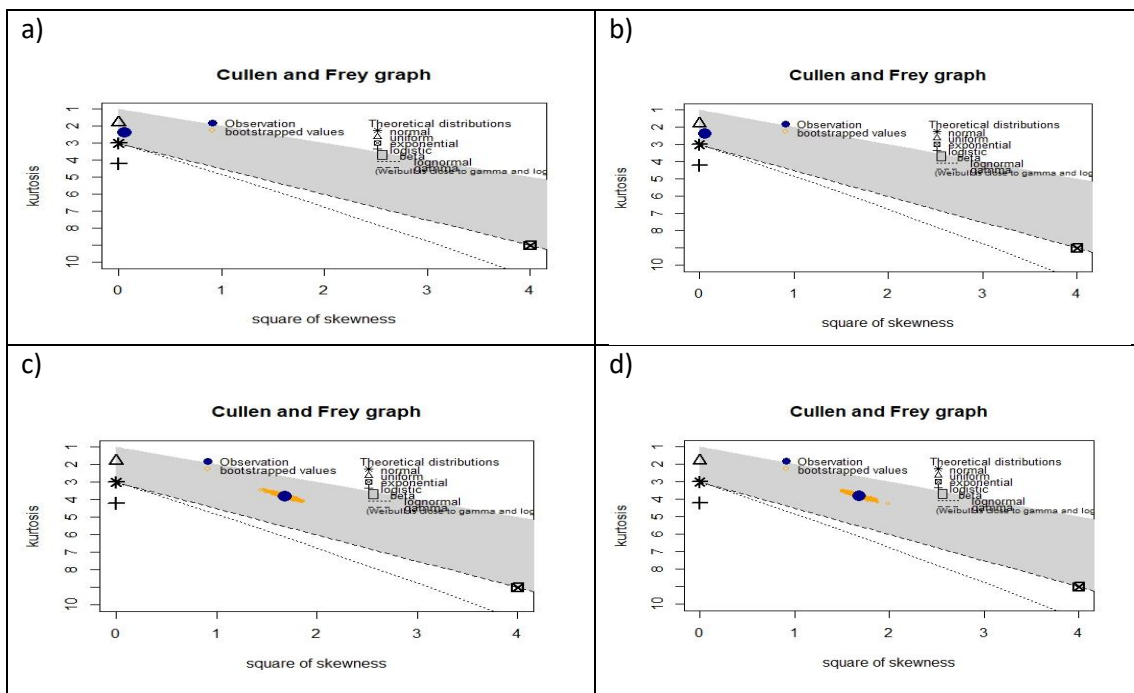
## Pearson\_1



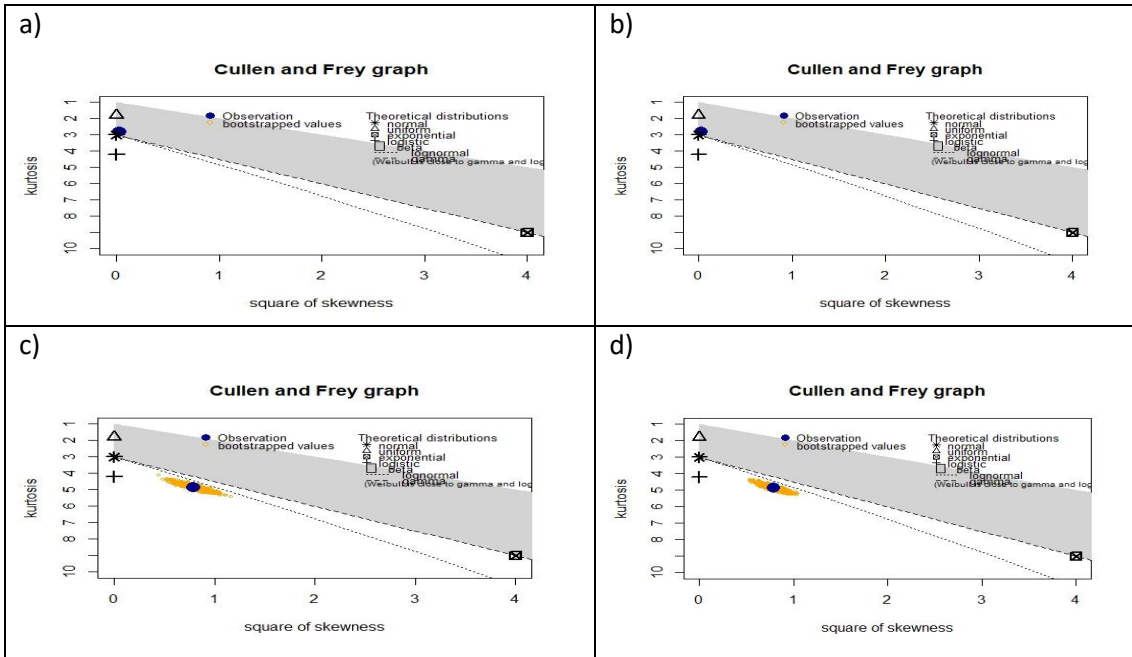
## Pearson\_2



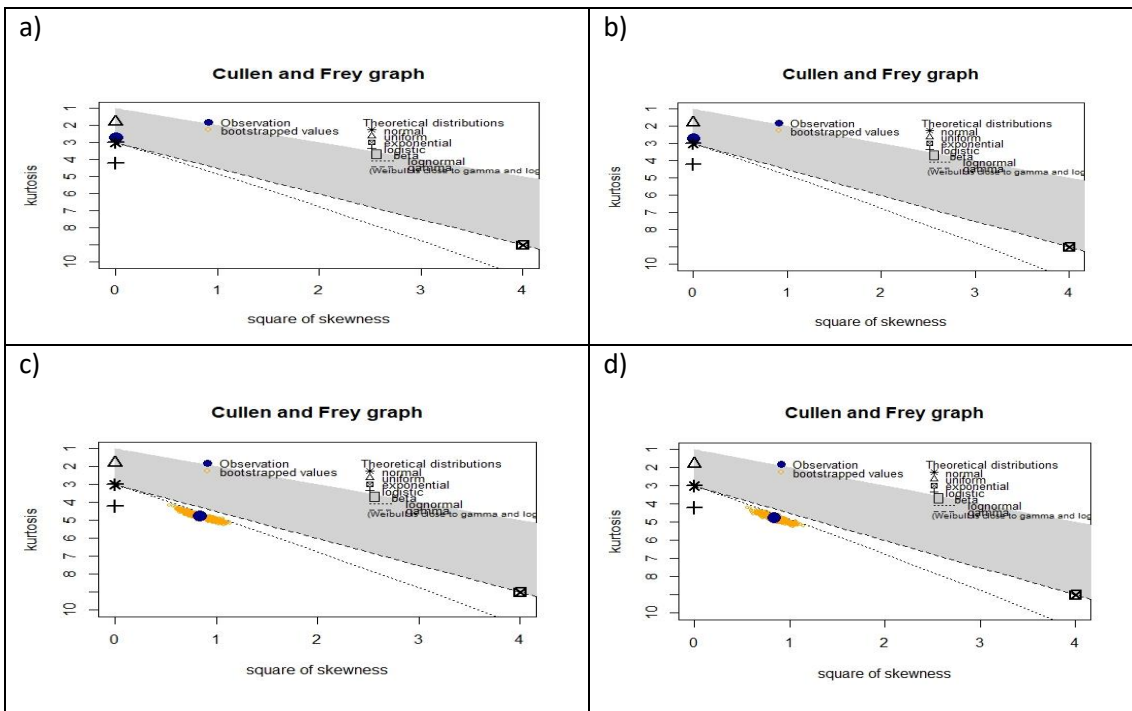
## Peirce\_1



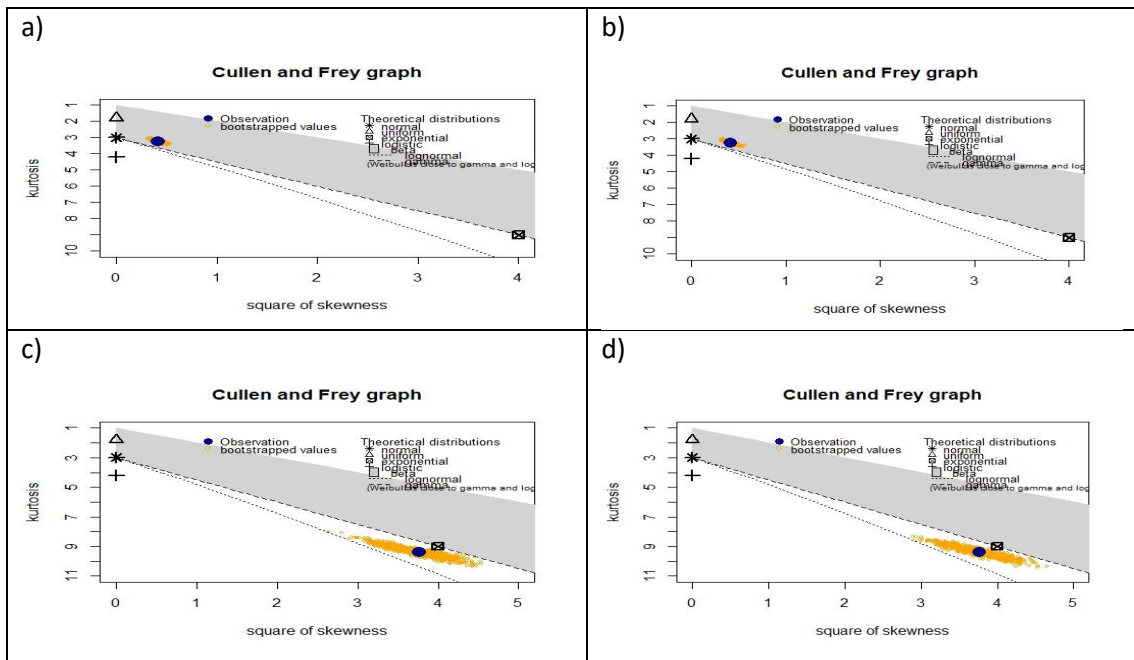
Peirce\_2



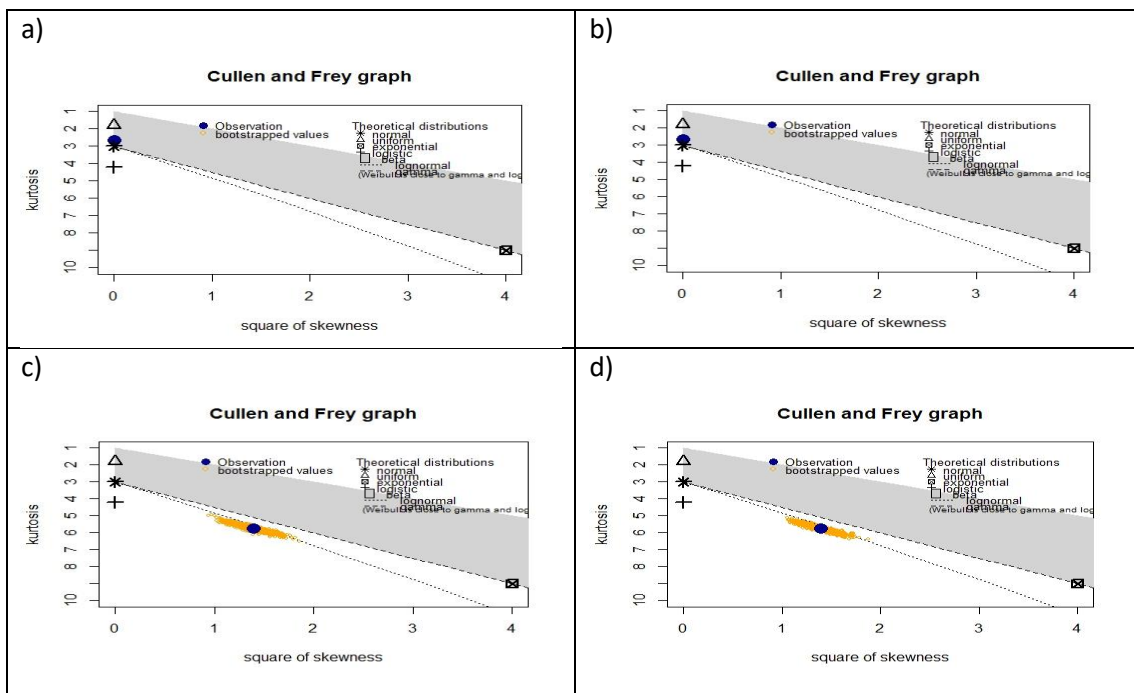
Peirce\_3



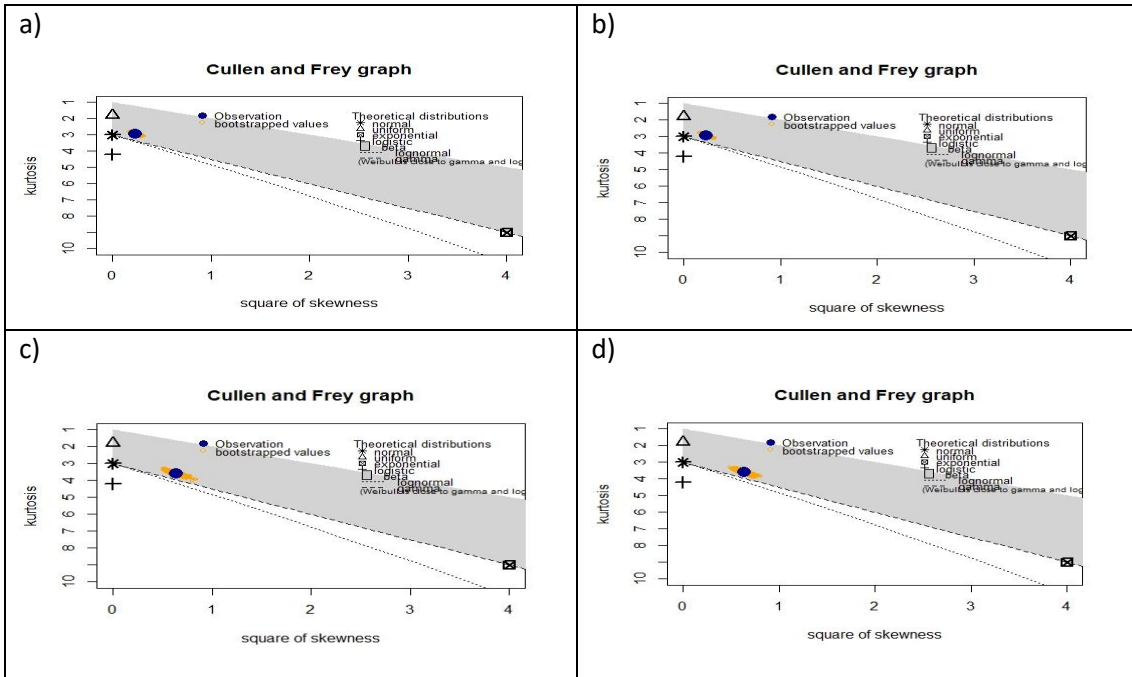
Rogers & Tanimoto



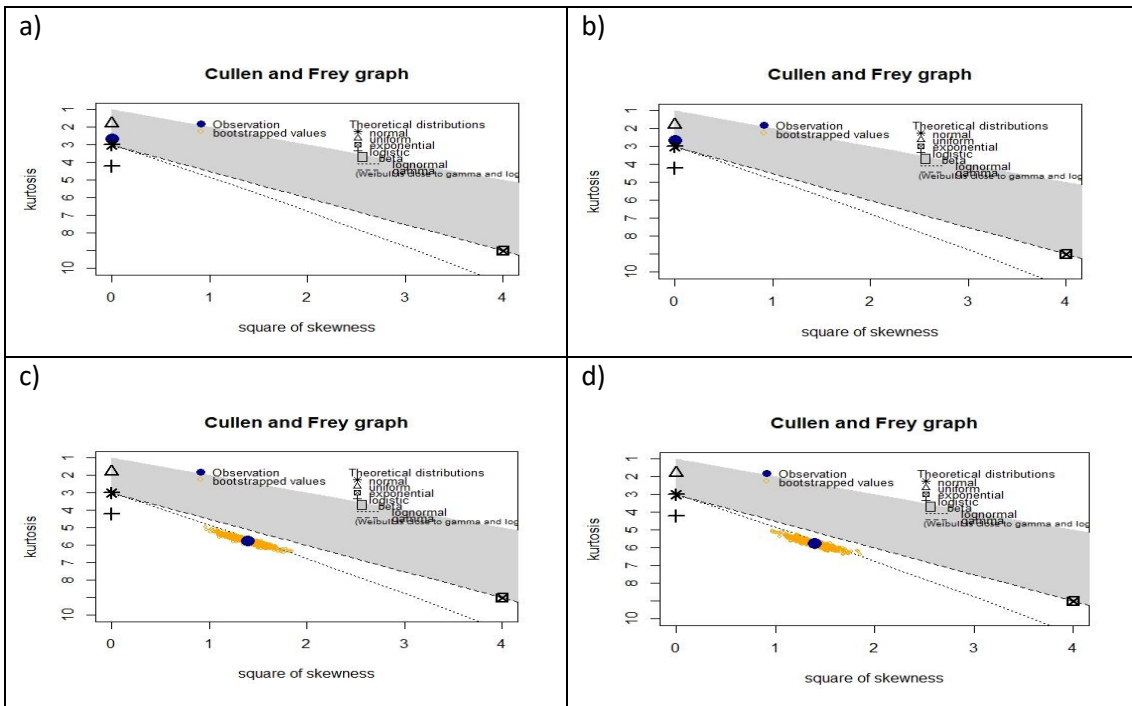
Rogot & Goldberg



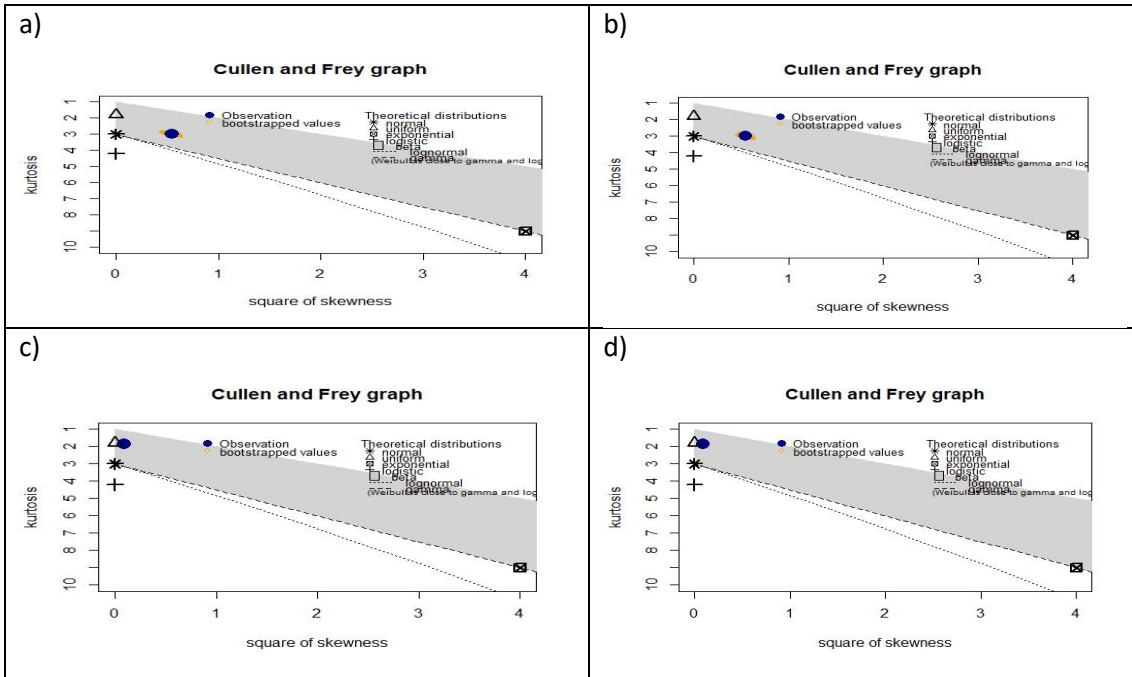
Russell & Rao



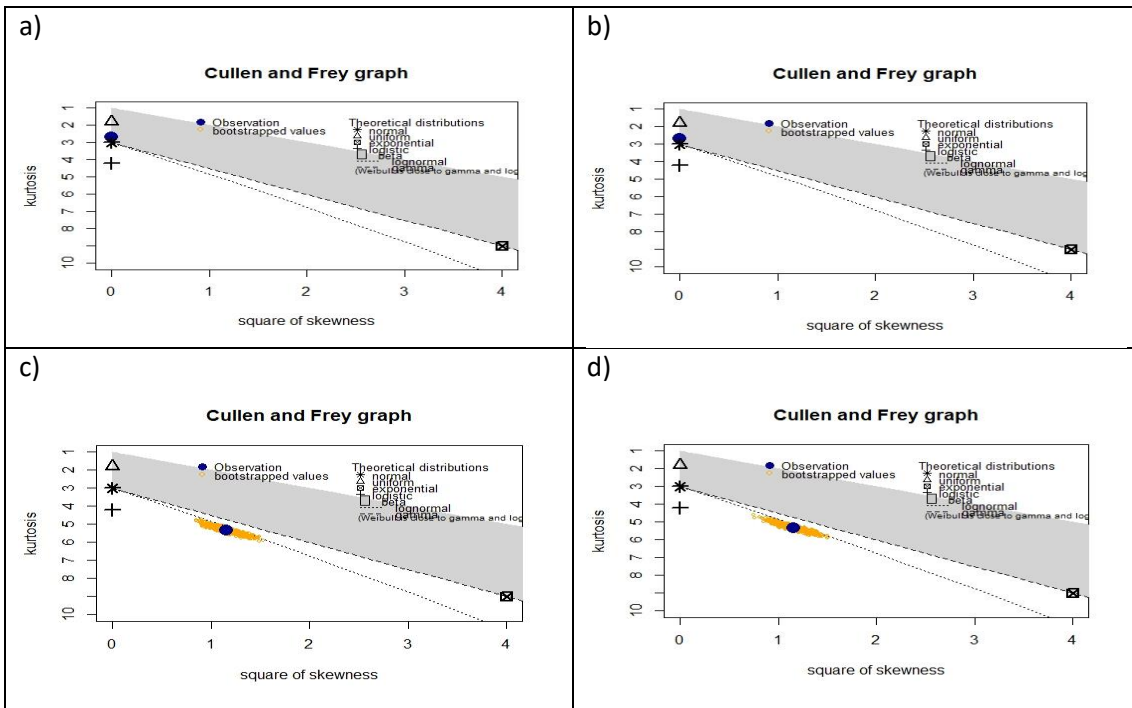
Scott



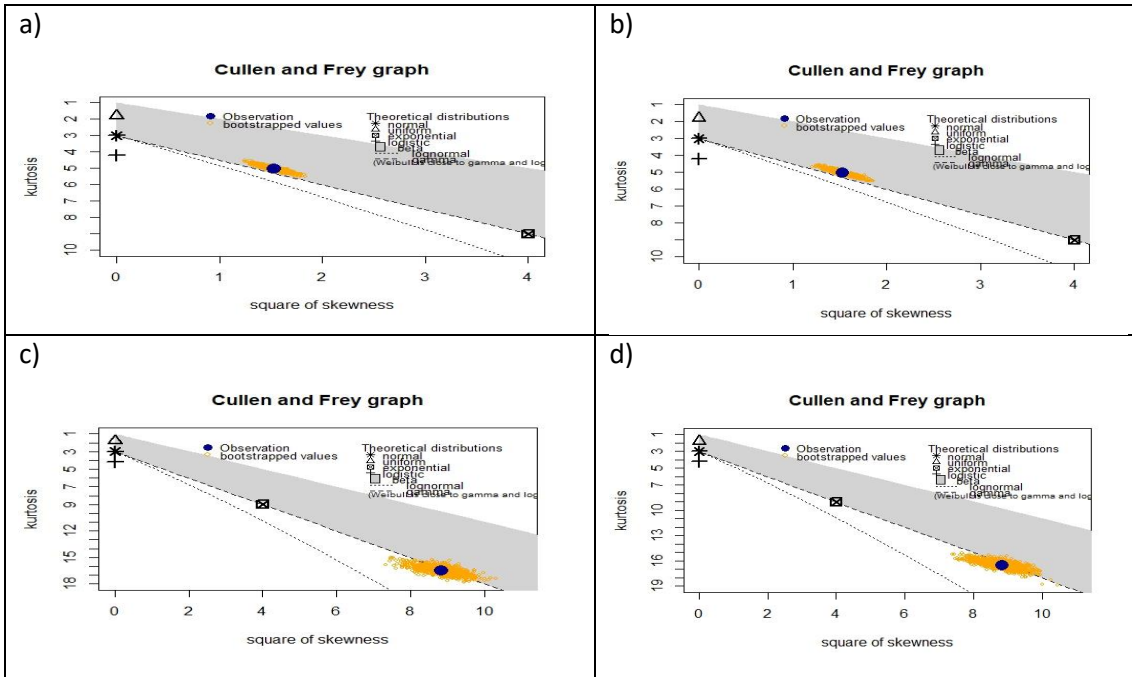
### Simpson



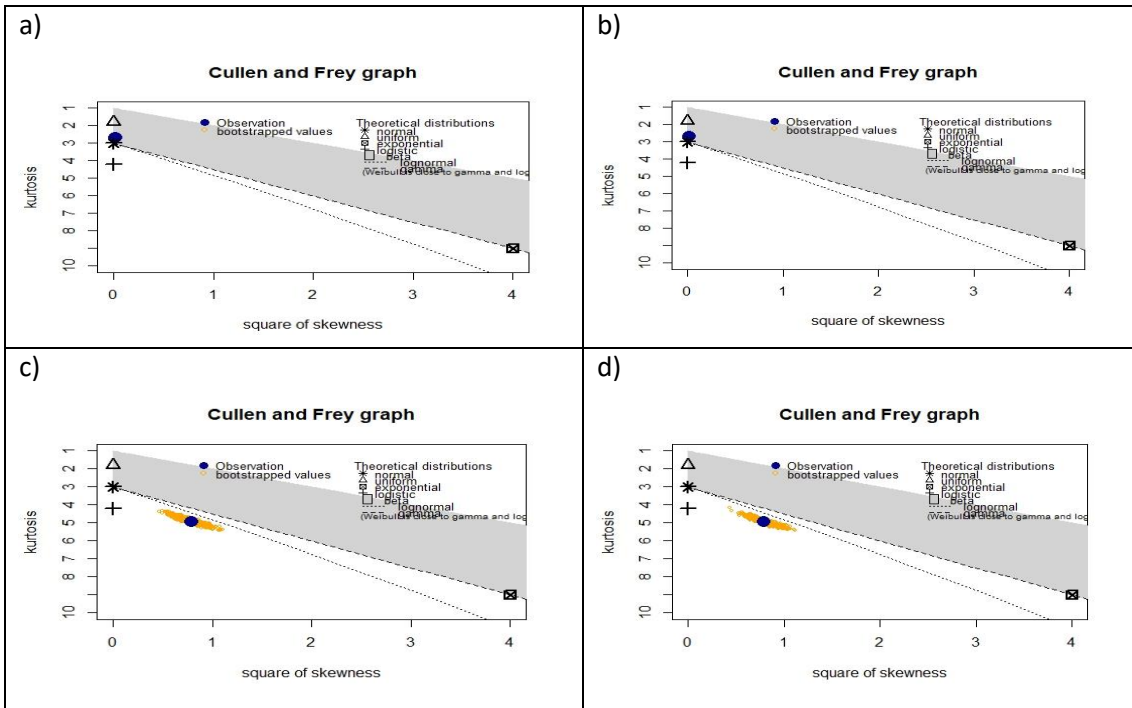
### Sokal & Michener



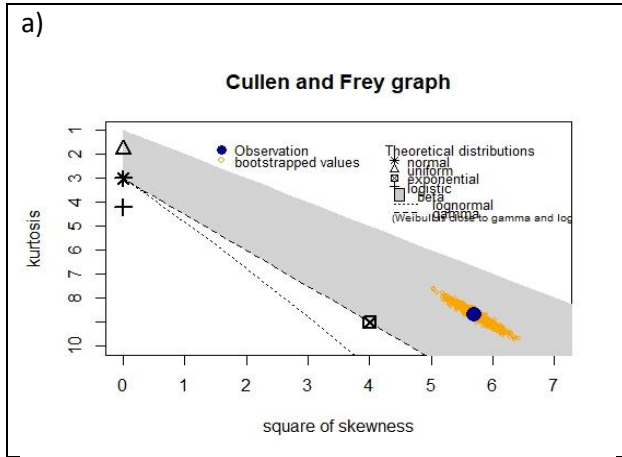
Sokal & Sneath\_1



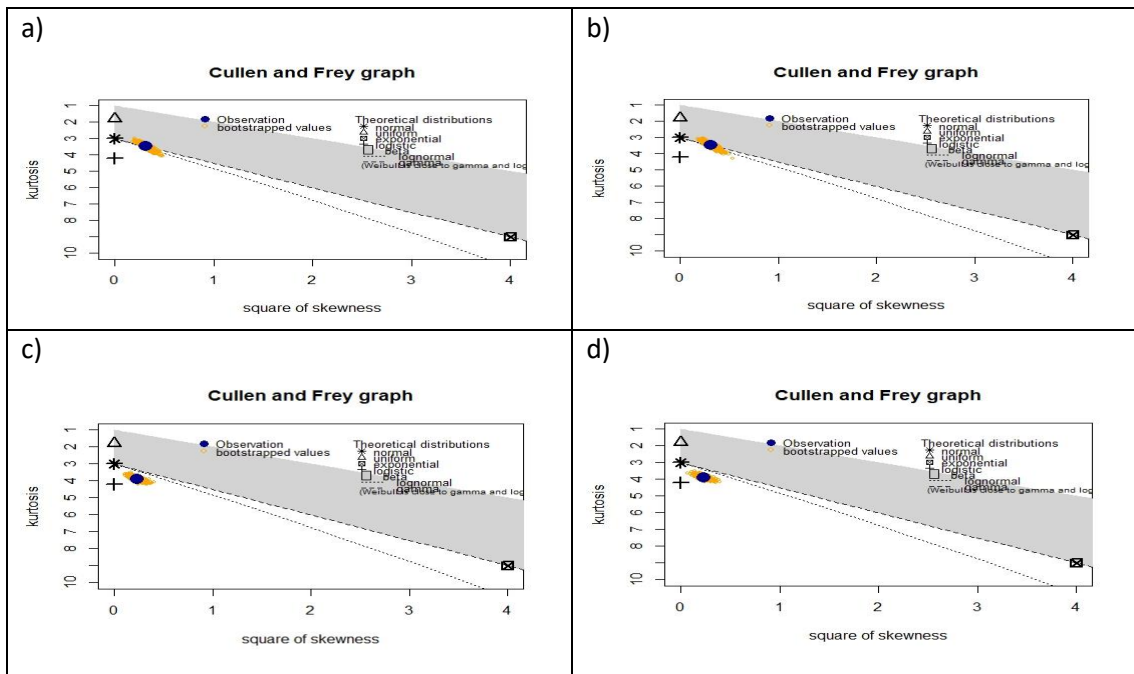
Sokal & Sneath\_2



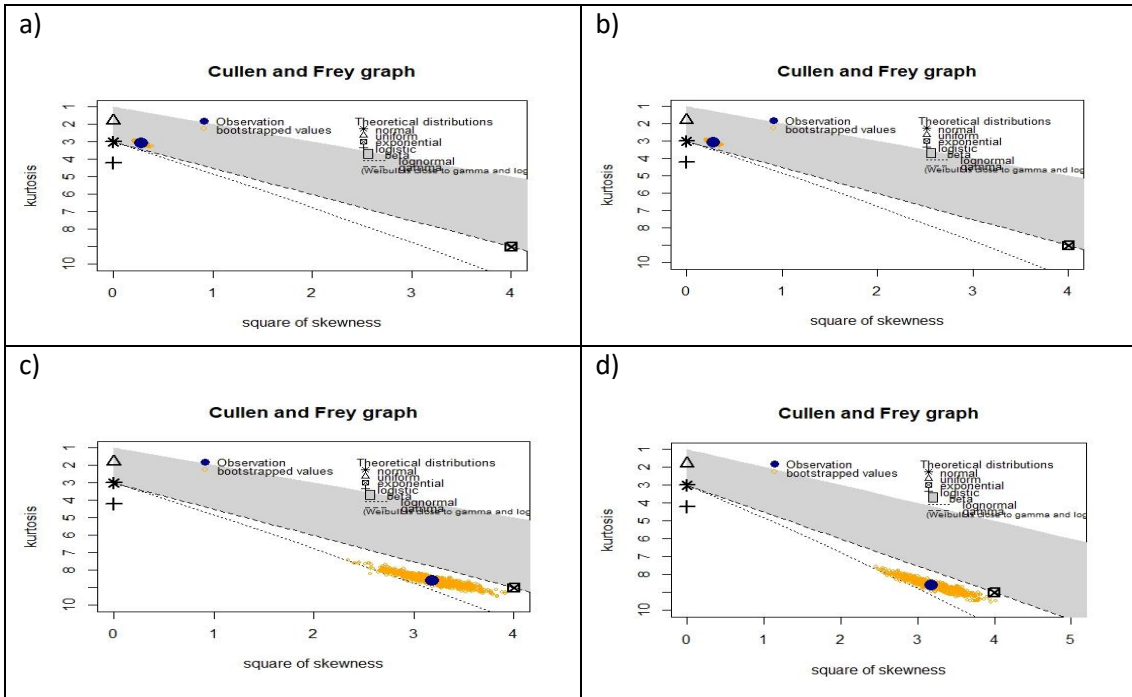
Sokal & Sneath\_3



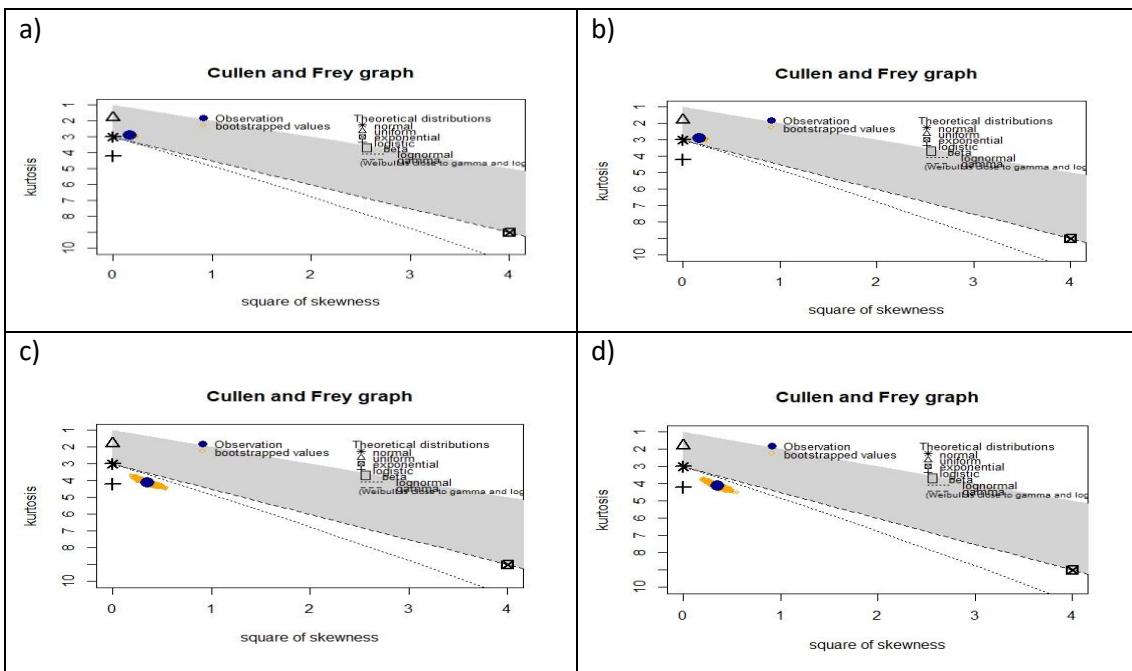
Sokal & Sneath\_4



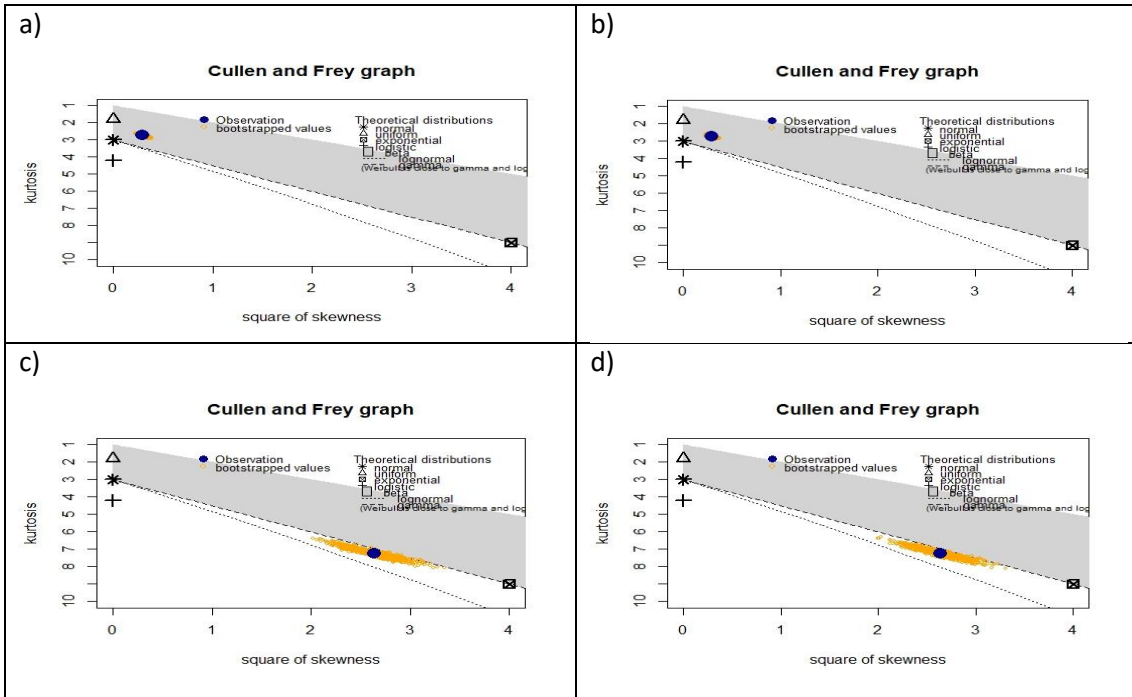
Sokal & Sneath\_5



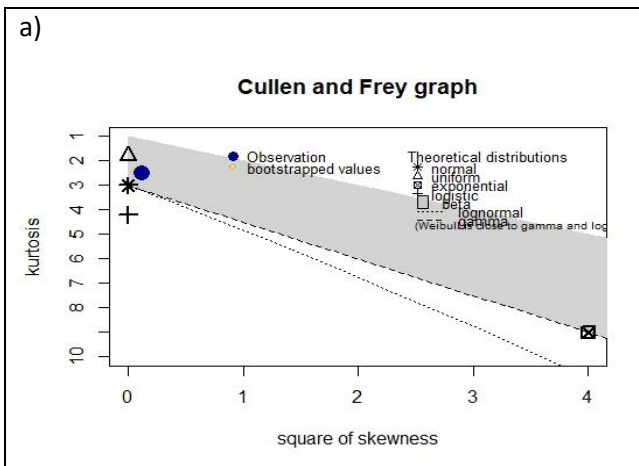
Sorensen



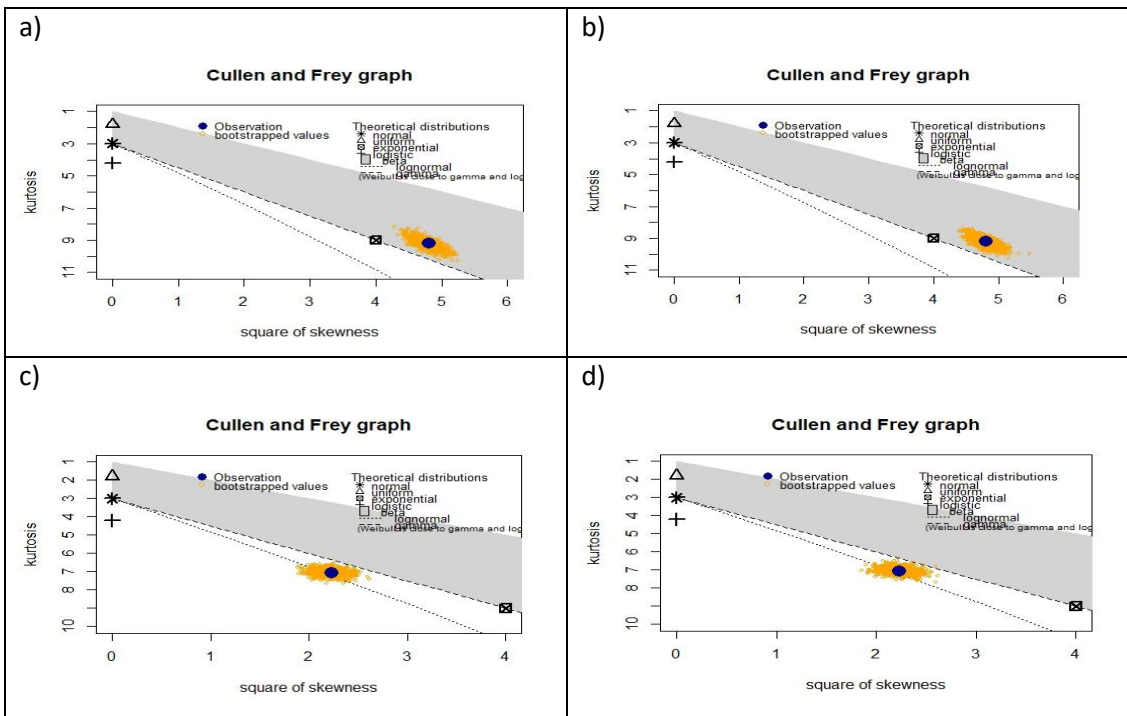
## Sorgenfrei



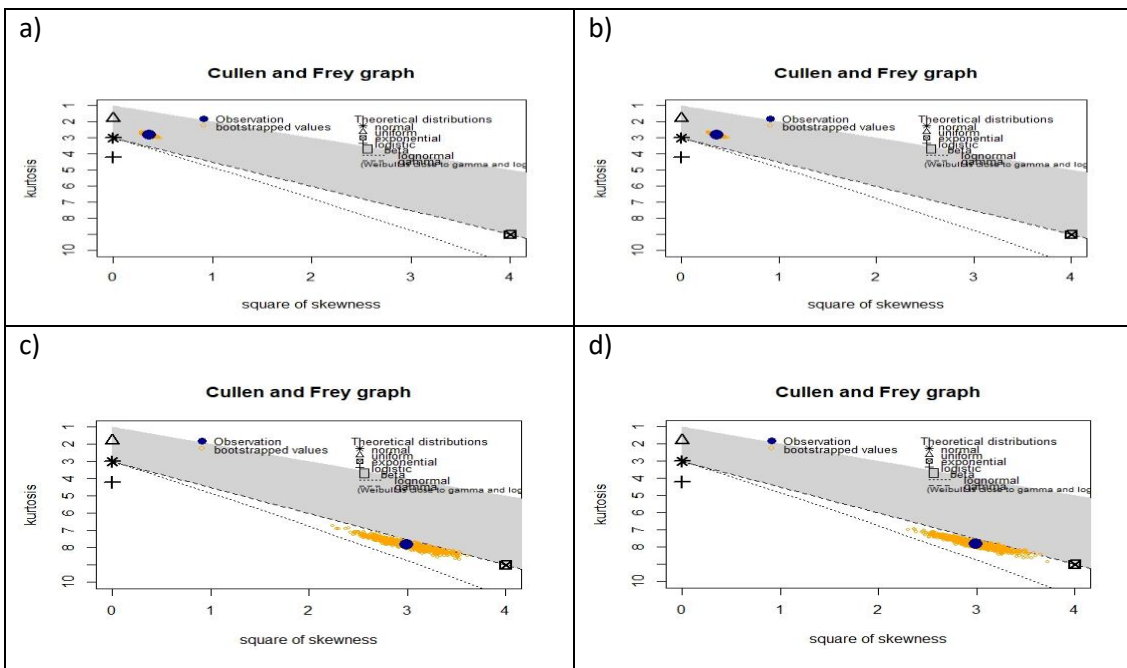
## Stiles



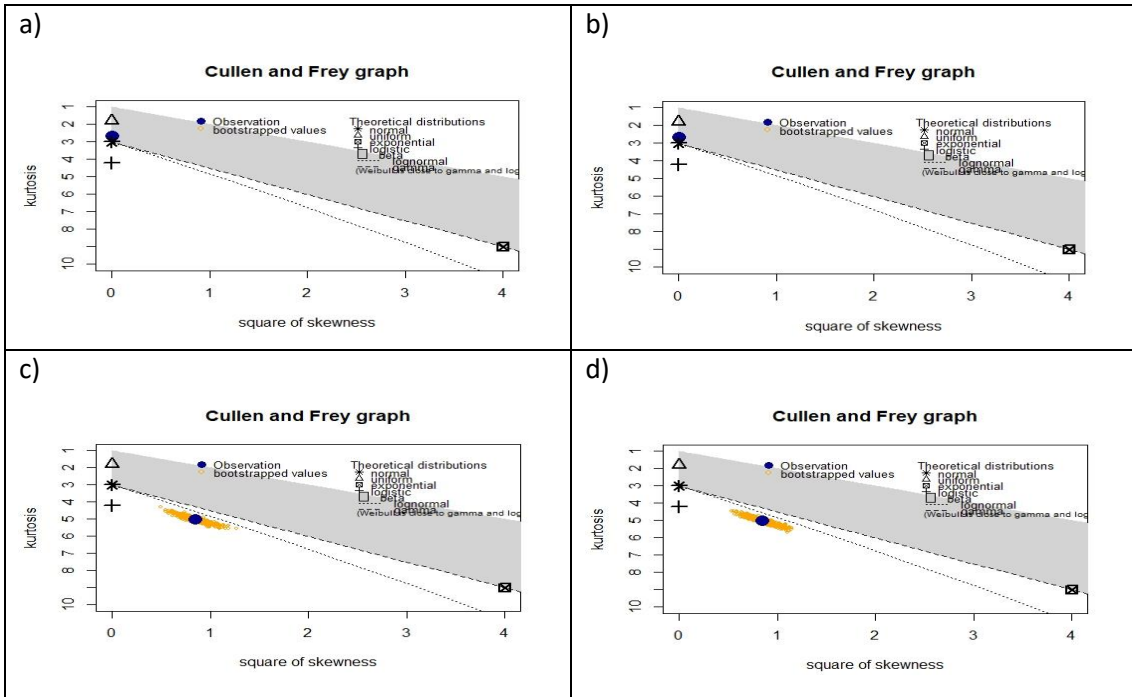
Tarwid



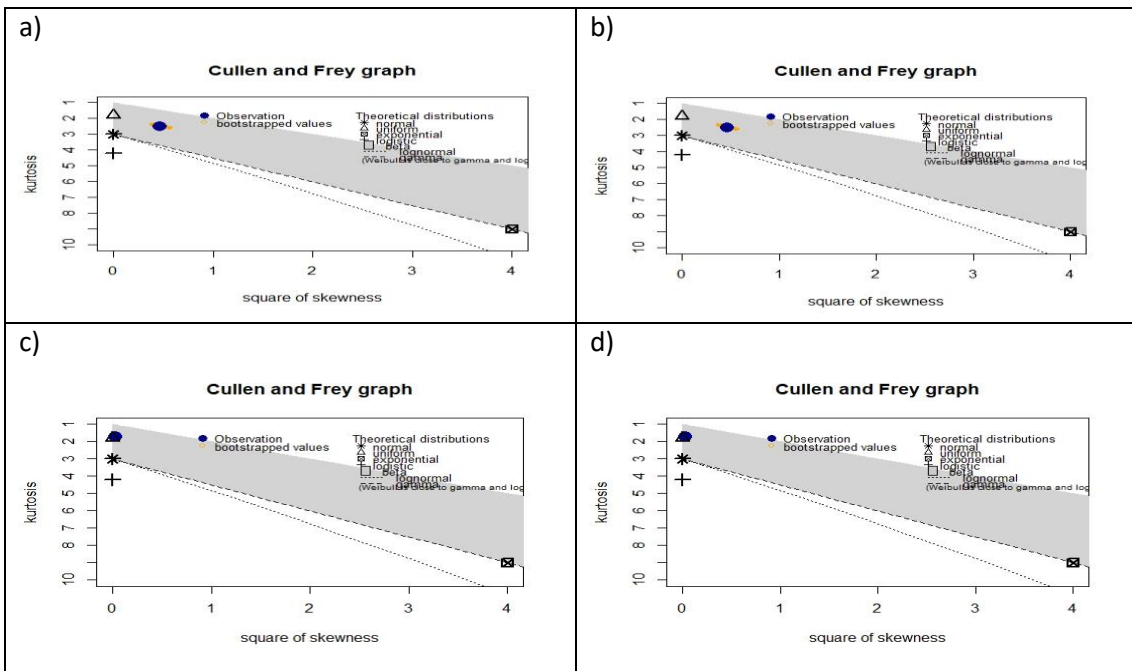
Tschuproff



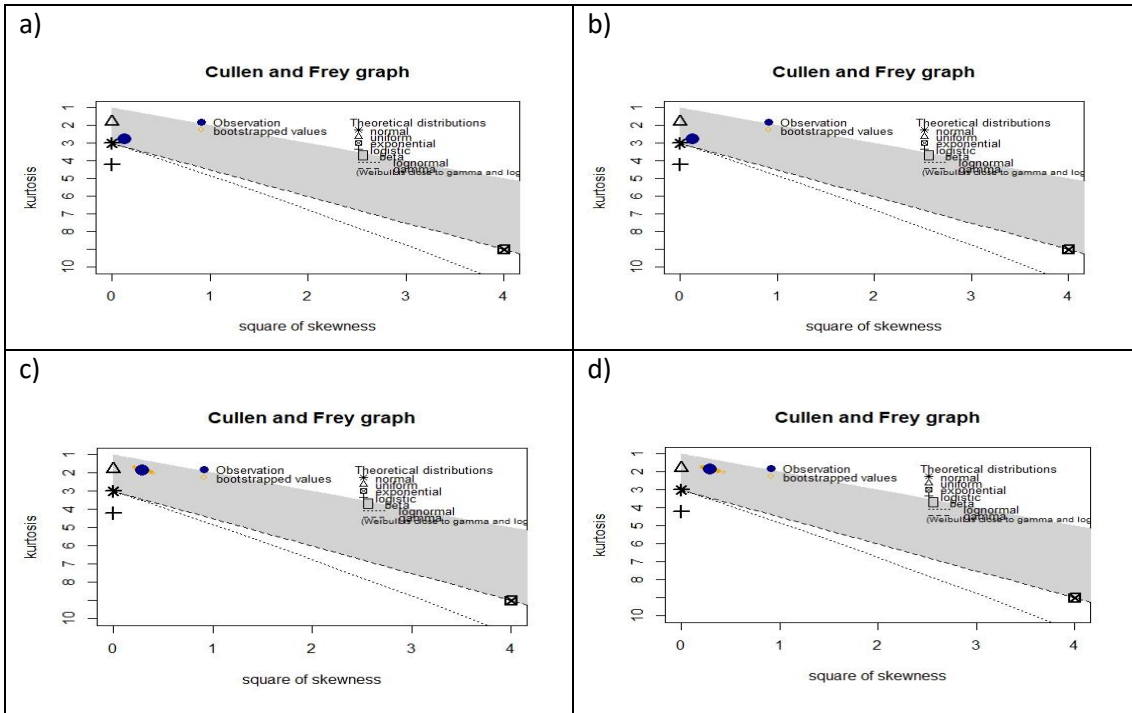
### Yule\_1



### Yule\_2



Yule\_3





## ANEXO 8: PROGRAMAS DE COMPUTACIÓN

### Programa Propiedad 2: Maximal y minimal

#### Prog.107

```
/* prog107 similitud Jaccard */
/* calcula la prop. 2: Maximal y Minimal de indices de sim. y dis. */

proc iml;
ja = j(100,100,0);
dja = j(100,100,0);
ddja = j(100,100,0);
dddja = j(100,100,0);
can = 1;
kat=0;
qat1=0;
qat2=0;
qat3=0;
indet=0;
nz=10;

/* setea los parametros */
/* n=tamaño del vector */
/* p=10 vectores */
/* p1=prob.1 p2=prob.2 ... p10=prob.10 */
n = 100;
p = 10;
z = j(p,n,0);
p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;

/* genera 100 vectores binomiales */
do h=1 to n by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
```

```
z[8,g]=z8;  
z[9,g]=z9;  
z[10,g]=z10;
```

```
seed=h;  
z1 = ranbin (seed, 1, p1);  
z2 = ranbin (seed, 1, p2);  
z3 = ranbin (seed, 1, p1);  
z4 = ranbin (seed, 1, p2);  
z5 = ranbin (seed, 1, p1);  
z6 = ranbin (seed, 1, p2);  
z7 = ranbin (seed, 1, p1);  
z8 = ranbin (seed, 1, p8);  
z9 = ranbin (seed, 1, p9);  
z10 = ranbin (seed, 1, p10);  
z[1,g+1]=z1;  
z[2,g+1]=z2;  
z[3,g+1]=z3;  
z[4,g+1]=z4;  
z[5,g+1]=z5;  
z[6,g+1]=z6;  
z[7,g+1]=z7;  
z[8,g+1]=z8;  
z[9,g+1]=z9;  
z[10,g+1]=z10;
```

```
seed=h;  
z1 = ranbin (seed, 1, p3);  
z2 = ranbin (seed, 1, p3);  
z3 = ranbin (seed, 1, p3);  
z4 = ranbin (seed, 1, p3);  
z5 = ranbin (seed, 1, p3);  
z6 = ranbin (seed, 1, p6);  
z7 = ranbin (seed, 1, p6);  
z8 = ranbin (seed, 1, p6);  
z9 = ranbin (seed, 1, p6);  
z10 = ranbin (seed, 1, p6);  
z[1,g+2]=z1;  
z[2,g+2]=z2;  
z[3,g+2]=z3;  
z[4,g+2]=z4;  
z[5,g+2]=z5;  
z[6,g+2]=z6;  
z[7,g+2]=z7;  
z[8,g+2]=z8;  
z[9,g+2]=z9;  
z[10,g+2]=z10;
```

```
seed=h;  
z1 = ranbin (seed, 1, p1);  
z2 = ranbin (seed, 1, p2);  
z3 = ranbin (seed, 1, p3);  
z4 = ranbin (seed, 1, p9);  
z5 = ranbin (seed, 1, p10);  
z6 = ranbin (seed, 1, p9);  
z7 = ranbin (seed, 1, p10);  
z8 = ranbin (seed, 1, p9);  
z9 = ranbin (seed, 1, p10);  
z10 = ranbin (seed, 1, p9);  
z[1,g+3]=z1;  
z[2,g+3]=z2;
```

```

z[3,g+3]=z3;
z[4,g+3]=z4;
z[5,g+3]=z5;
z[6,g+3]=z6;
z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to n;
  do k = i to n;
    a=0;
    b=0;
    c=0;
    d=0;
    aprima=0;
    do o = 1 to p;
      if (z[o,i]<z[o,k]) then c=c+abs(z[o,i]-z[o,k]);
      if (z[o,i]>z[o,k]) then b=b+abs(z[o,i]-z[o,k]);
      a=a+z[o,i]*z[o,k];
      d=d+(1-z[o,i])*(1-z[o,k]);
    end;
    if (a+b+c = 0) then indet = indet + 1;
    if (a+b+c = 0) then tja=0;
    if (a+b+c > 0) then tja=(a)/(a+b+c);
    tdja=1-tja;
    tddja=sqrt(tdja);
    tdddja=sqrt(2*tdja);
    ja[i,k] = tja;
    dja[i,k] = tdja;
    ddja[i,k] = tddja;
    dddja[i,k] = tdddja;
    can=can+1;
  end;
end;

do i=1 to n;

```

```
do j=i+1 to n;
if (ja[i,i]<ja[i,j]) then kat=kat+1;
if (dja[i,i]>dja[i,j]) then qat1=qat1+1;
if (ddja[i,i]>ddja[i,j]) then qat2=qat2+1;
if (dddja[i,i]>dddja[i,j]) then qat3=qat3+1;
end;
end;

print ja dja ddja dddja can kat qat1 qat2 qat3 indet;
```

## Programa Propiedad 6: Simetria Prog.108

```
/* prog108 similitud Jaccard */
/* calcula la prop. 6, Simetria de indices de sim. y dis. */

proc iml;
ja = j(100,100,0);
dja = j(100,100,0);
ddja = j(100,100,0);
dddja = j(100,100,0);
can = 0;
kat=0;
pat1=0;
pat2=0;
pat3=0;
indet=0;
nz=10;

/* setea los parametros */
/* n=tamaño del vector */
/* p=10 vectores */
/* p1=prob.1 p2=prob.2 ... p10=prob.10 */
n = 100;
p = 10;
z = j(p,n,0);
p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;

/* genera 100 vectores binomiales */
do h=1 to n by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
z[8,g]=z8;
z[9,g]=z9;
z[10,g]=z10;
```

```

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p2);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p2);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g+1]=z1;
z[2,g+1]=z2;
z[3,g+1]=z3;
z[4,g+1]=z4;
z[5,g+1]=z5;
z[6,g+1]=z6;
z[7,g+1]=z7;
z[8,g+1]=z8;
z[9,g+1]=z9;
z[10,g+1]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p3);
z2 = ranbin (seed, 1, p3);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p3);
z5 = ranbin (seed, 1, p3);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p6);
z8 = ranbin (seed, 1, p6);
z9 = ranbin (seed, 1, p6);
z10 = ranbin (seed, 1, p6);
z[1,g+2]=z1;
z[2,g+2]=z2;
z[3,g+2]=z3;
z[4,g+2]=z4;
z[5,g+2]=z5;
z[6,g+2]=z6;
z[7,g+2]=z7;
z[8,g+2]=z8;
z[9,g+2]=z9;
z[10,g+2]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p10);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p10);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p10);
z10 = ranbin (seed, 1, p9);
z[1,g+3]=z1;
z[2,g+3]=z2;
z[3,g+3]=z3;
z[4,g+3]=z4;
z[5,g+3]=z5;
z[6,g+3]=z6;

```

```

z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to n;
  do k = 1 to n;
    a=0;
    b=0;
    c=0;
    d=0;
    aprima=0;
    do o = 1 to p;
      if (z[o,i]<z[o,k]) then c=c+abs(z[o,i]-z[o,k]);
      if (z[o,i]>z[o,k]) then b=b+abs(z[o,i]-z[o,k]);
      a=a+z[o,i]*z[o,k];
      d=d+(1-z[o,i])*(1-z[o,k]);
    end;
    nume=a;
    deno=a+b+c;
    if (deno = 0) then tja=0;
    else tja=nume/deno;
    tdja=(1-tja);
    tddja=sqrt(tdja);
    tdddja=sqrt(2*tddja);
    ja[i,k] = tja;
    dja[i,k] = tdja;
    ddja[i,k] = tddja;
    dddja[i,k] = tdddja;
    can=can+1;
  end;
end;

do i=1 to n;
  do j=i+1 to n;
    if (ja[i,j] ^= ja[j,i]) then kat=kat+1;
    if (dja[i,j] ^= dja[j,i]) then pat1=pat1+1;
  end;
end;

```

```
    if (ddja[i,j] ^= ddja[j,i]) then pat2=pat2+1;
    if (dddja[i,j] ^= dddja[j,i]) then pat3=pat3+1;
end;
print ja dja ddja dddja can kat pat1 pat2 pat3;
```

## Programas Propiedad 7: Monotonicidad respecto a la estadística $\chi^2$

### Prog.106

```
/* prog106 similitud Jaccard */
/* calcula la prop.7, Monotonicidad de indices de sim. */

proc iml;
ja = j(4950,2,0);
can = 1;
nz=10;

/* setea los parámetros, n=tamaño del vector, p=10 vectores */
/* p1=prob.1 p2=prob.2 ... p10=prob.10 */
n = 100;
p = 10;
z = j(p,n,0);
p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;

/* genera 100 vectores binomiales */
do h=1 to 100 by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
z[8,g]=z8;
z[9,g]=z9;
z[10,g]=z10;

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p2);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p2);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
```

```

z10 = ranbin (seed, 1, p10);
z[1,g+1]=z1;
z[2,g+1]=z2;
z[3,g+1]=z3;
z[4,g+1]=z4;
z[5,g+1]=z5;
z[6,g+1]=z6;
z[7,g+1]=z7;
z[8,g+1]=z8;
z[9,g+1]=z9;
z[10,g+1]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p3);
z2 = ranbin (seed, 1, p3);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p3);
z5 = ranbin (seed, 1, p3);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p6);
z8 = ranbin (seed, 1, p6);
z9 = ranbin (seed, 1, p6);
z10 = ranbin (seed, 1, p6);
z[1,g+2]=z1;
z[2,g+2]=z2;
z[3,g+2]=z3;
z[4,g+2]=z4;
z[5,g+2]=z5;
z[6,g+2]=z6;
z[7,g+2]=z7;
z[8,g+2]=z8;
z[9,g+2]=z9;
z[10,g+2]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p10);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p10);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p10);
z10 = ranbin (seed, 1, p9);
z[1,g+3]=z1;
z[2,g+3]=z2;
z[3,g+3]=z3;
z[4,g+3]=z4;
z[5,g+3]=z5;
z[6,g+3]=z6;
z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);

```

```

z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to 100;
  do k = i+1 to 100;
    a=0;
    b=0;
    c=0;
    d=0;
    aprima=0;
    do o = 1 to 10;
      if (z[o,i]<z[o,k]) then c=c+abs(z[o,i]-z[o,k]);
      if (z[o,i]>z[o,k]) then b=b+abs(z[o,i]-z[o,k]);
      a=a+z[o,i]*z[o,k];
      d=d+(1-z[o,i])*(1-z[o,k]);
    end;
    ppp=a+b+c;
    if (ppp = 0) then tja=0;
    else tja=a/ppp;
    ene=a+b+c+d;
    nume=a*d-b*c;
    chi=(ene*(nume*nume))/((a+b)*(c+d)*(a+c)*(b+d));
    chi2=sqrt(chi);
    ja[can,1] = tja;
    ja[can,2] = chi2;
    can=can+1;
  end;
end;

print ja;
quit;

```

## Prog.99

```

/* prog.99 */
/* Calcula correlaciones de Spearman y Kendall */
/* entre los valores de Jaccard y Chi cuadrado */

data pato;
input ja chi2;
cards;
0 0.3513642
0 0.5270463
0 0.5270463

```

```
0 0.5270463
0 0.5270463
0 0.5270463
```

```
.....
```

```
0.8 2.5819889
0.8 2.5819889
0.8 2.5819889
0.8 2.5819889
0.8 2.5819889
0.8 2.5819889
```

```
.....
```

```
1 3.1622777
1 3.1622777
1 3.1622777
1 3.1622777
1 3.1622777
1 3.1622777
```

```
;
```

```
proc corr spearman data = pato;
run;
```

```
proc corr kendall data=pato;
run;
quit;
```

## Programas Propiedad 8: Linealidad

### Prog.200

```
/* prog200Jaccard */
/* propiedad 8, linealidad de coef. de similitud */
/* calcula el indice de Jaccard (tja) */

proc iml;

AN=0;
BE=40;
CE=40;
DE=20;

/* Indice de similitud de Jaccard (tja con a, b, c, d) */
do i= 1 to 4;
    a=AN;
    b=BE;
    c=CE;
    d=DE;
    tja=(a)/(a+b+c);

/* Indice de similitud de Jaccard (tja con a+1, b-1, c-1, d+1) */
    a=AN+1;
    b=BE-1;
    c=CE-1;
    d=DE+1;
    tja1=(a)/(a+b+c);

/* Indice de similitud de Jaccard (tja con a+2, b-2, c-2, d+2) */
    a=AN+2;
    b=BE-2;
    c=CE-2;
    d=DE+2;
    tja2=(a)/(a+b+c);
print tja tja1 tja2;

/* prueba la propiedad de linealidad */
prima=tja1-tja;
seconda=tja2-tja1;
if (prima=seconda) then lineal=0;
if (prima<seconda) then lineal=-1;
if (prima>seconda) then lineal=1;
print prima seconda lineal;

AN=AN+1;
BE=BE-1;
CE=CE-1;
DE=DE+1;
end;
quit;
```

### Prog.410

```
/* prog410Jaccard */
/* propiedad de resolucio, no linealidad y CV de coef. de similitud
*/
/* calcula el indice de Jaccard (tja) */

proc iml;
```

```

tja=j(27,1,0);
tdja=j(27,1,0);
tddja=j(27,1,0);
tdddja=j(27,1,0);
la=j(24,1,0);
pa=j(24,1,0);
ka=j(24,1,0);
ro=j(24,1,0);

/* Serie 1 */
AN=45;
BE=5;
CE=50;
DE=50;

/* Indice de similitud de Jaccard (tja con a, b, c, d) */
do i= 1 to 9;
    a=AN;
    b=BE;
    c=CE;
    d=DE;
    sim=a/(a+b+c);
    dis=1-sim;
    tja[i,1]=sim;
    tdja[i,1]=dis;
    tddja[i,1]=sqrt(dis);
    tdddja[i,1]=sqrt(2*dis);

    AN=AN-5;
    BE=BE+5;
    CE=CE;
    DE=DE;
end;

/* Serie 2 */
AN=90;
BE=10;
CE=0;
DE=50;

/* Indice de similitud de Jaccard (tja con a, b, c, d) */
do i= 10 to 18;
    a=AN;
    b=BE;
    c=CE;
    d=DE;
    sim=a/(a+b+c);
    dis=1-sim;
    tja[i,1]=sim;
    tdja[i,1]=dis;
    tddja[i,1]=sqrt(dis);
    tdddja[i,1]=sqrt(2*dis);

    AN=AN-10;
    BE=BE+10;
    CE=CE;
    DE=DE;
end;

/* Serie 3 */
AN=90;

```

```

BE=5;
CE=5;
DE=50;

/* Indice de similitud de Jaccard (tja con a, b, c, d) */
do i= 19 to 27;
    a=AN;
    b=BE;
    c=CE;
    d=DE;
    sim=a/(a+b+c);
    dis=1-sim;
    tja[i,1]=sim;
    tdja[i,1]=dis;
    tddja[i,1]=sqrt(dis);
    tdddja[i,1]=sqrt(2*dis);

    AN=AN-10;
    BE=BE+5;
    CE=CE+5;
    DE=DE;
end;
print tja tdja tddja tdddja;

/* primer vector */
do i= 1 to 8;
    j= i+1;
    la[i,1]=abs(tja[j,1]-tja[i,1]);
    pa[i,1]=abs(tdja[j,1]-tdja[i,1]);
    ka[i,1]=abs(tddja[j,1]-tddja[i,1]);
    ro[i,1]=abs(tdddja[j,1]-tdddja[i,1]);
end;

/* segundo vector */
do i= 10 to 17;
    j= i+1;
    k= i-1;
    la[k,1]=abs(tja[j,1]-tja[i,1]);
    pa[k,1]=abs(tdja[j,1]-tdja[i,1]);
    ka[k,1]=abs(tddja[j,1]-tddja[i,1]);
    ro[k,1]=abs(tdddja[j,1]-tdddja[i,1]);
end;

/* tercer vector */
do i= 19 to 26;
    j= i+1;
    m= i-2;
    la[m,1]=abs(tja[j,1]-tja[i,1]);
    pa[m,1]=abs(tdja[j,1]-tdja[i,1]);
    ka[m,1]=abs(tddja[j,1]-tddja[i,1]);
    ro[m,1]=abs(tdddja[j,1]-tdddja[i,1]);
end;
print la pa ka ro;

/* calcula linealidad, resolucion y CV */
sumtja=0;
sumtja2=0;
sumtdja=0;
sumtdja2=0;
sumtddja=0;
sumtddja2=0;

```

```

sumtdddja=0;
sumtdddja2=0;
do i = 1 to 24;
sumtja=sumtja+la[i,1];
sumtja2=sumtja2+(la[i,1])##2;
sumtdja=sumtdja+pa[i,1];
sumtdja2=sumtdja2+(pa[i,1])##2;
sumtddja=sumtddja+ka[i,1];
sumtddja2=sumtddja2+(ka[i,1])##2;
sumtdddja=sumtdddja+ro[i,1];
sumtdddja2=sumtdddja2+(ro[i,1])##2;
end;

medtja=sumtja/24;
vartja=(sumtja2-(24*medtja*medtja))/23;
destja=sqrt(vartja);
cvtja=(destja/medtja)*100;

medtdja=sumtdja/24;
vartdja=(sumtdja2-(24*medtdja*medtdja))/23;
destdja=sqrt(vartdja);
cvttdja=(destdja/medtdja)*100;

medtddja=sumtddja/24;
vartddja=(sumtddja2-(24*medtddja*medtddja))/23;
destddja=sqrt(vartddja);
cvtddja=(destddja/medtddja)*100;

medtdddja=sumtdddja/24;
vartdddja=(sumtdddja2-(24*medtdddja*medtdddja))/23;
destdddja=sqrt(vartdddja);
cvtdddja=(destdddja/medtdddja)*100;

print medtja destja cvtja;
print medtdja destdja cvttdja;
print medtddja destddja cvtddja;
print medtdddja destdddja cvtdddja;

quit;

```

## Programa Propiedad 9: Insensibilidad al valor de a=0 Prog.100

```
/* prog100Jaccard */
/* propiedad 9, insensibilidad al valor de a=0 */
/* calcula el indice de Jaccard (tja) */

proc iml;

tja=j(40,3,0);
prima=j(40,1,0);
seconda=j(40,1,0);

AN=0;
BE=40;
CE=40;
DE=20;

/* Indice de similitud de Jaccard (tja con a, b, c, d) */
do i= 1 to 40;
    a=AN;
    b=BE;
    c=CE;
    d=DE;
    tja[i,1]=(a)/(a+b+c);

    a=AN+1;
    b=BE-1;
    c=CE-1;
    d=DE+1;
    tja[i,2]=(a)/(a+b+c);

    a=AN+2;
    b=BE-2;
    c=CE-2;
    d=DE+2;
    tja[i,3]=(a)/(a+b+c);

    prima[i]=tja[i,2]-tja[i,1];
    seconda[i]=tja[i,3]-tja[i,2];

    AN=AN+1;
    BE=BE-1;
    CE=CE-1;
    DE=DE+1;
end;

print tja prima seconda;
quit;
```

## Programa Propiedad 11: Distribuciones en el muestreo Prog.104

```
/* prog104 similitud Jaccard */
/* calcula la Prop.11: DISTRIBUCION de indices de sim. y dis. */

proc iml;
ja = j(4950,4,0);
can = 1;
nz=10;

/* setea los parametros n=tamaño del vector p=10 vectores */
/* p1=prob.1 p2=prob.2 ... p10=prob.10 */
n = 100;
p = 10;
z = j(p,n,0);
p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;

/* genera 100 vectores binomiales */
do h=1 to n by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
z[8,g]=z8;
z[9,g]=z9;
z[10,g]=z10;

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p2);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p2);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
```

```
z10 = ranbin (seed, 1, p10);
z[1,g+1]=z1;
z[2,g+1]=z2;
z[3,g+1]=z3;
z[4,g+1]=z4;
z[5,g+1]=z5;
z[6,g+1]=z6;
z[7,g+1]=z7;
z[8,g+1]=z8;
z[9,g+1]=z9;
z[10,g+1]=z10;
```

```
seed=h;
z1 = ranbin (seed, 1, p3);
z2 = ranbin (seed, 1, p3);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p3);
z5 = ranbin (seed, 1, p3);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p6);
z8 = ranbin (seed, 1, p6);
z9 = ranbin (seed, 1, p6);
z10 = ranbin (seed, 1, p6);
z[1,g+2]=z1;
z[2,g+2]=z2;
z[3,g+2]=z3;
z[4,g+2]=z4;
z[5,g+2]=z5;
z[6,g+2]=z6;
z[7,g+2]=z7;
z[8,g+2]=z8;
z[9,g+2]=z9;
z[10,g+2]=z10;
```

```
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p10);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p10);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p10);
z10 = ranbin (seed, 1, p9);
z[1,g+3]=z1;
z[2,g+3]=z2;
z[3,g+3]=z3;
z[4,g+3]=z4;
z[5,g+3]=z5;
z[6,g+3]=z6;
z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;
```

```
seed=20;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);
```

```

z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to 100;
  do k = i+1 to 100;
    a=0;
    b=0;
    c=0;
    d=0;
    do o = 1 to 10;
      if (z[o,i]<z[o,k]) then c=c+abs(z[o,i]-z[o,k]);
      if (z[o,i]>z[o,k]) then b=b+abs(z[o,i]-z[o,k]);
      a=a+z[o,i]*z[o,k];
      d=d+(1-z[o,i])*(1-z[o,k]);
    end;
    deno=a+b+c;
    if (deno = 0) then jac=0;
    else jac=a/deno;
    tja=jac;
    tdja=1-jac;
    tddja=sqrt(1-jac);
    tdddja=sqrt(2*(1-jac));
    ja[can,1] = tja;
    ja[can,2] = tdja;
    ja[can,3] = tddja;
    ja[can,4] = tdddja;
    can=can+1;
  end;
end;
print ja;
quit;

```

## Programa Propiedad 12: Insesgamiento Prog.211

```
/* prog211 similitud Jaccard */
/* calcula la Prop.12, insesgamiento, simula datos multinomiales */
/* selecciona 1000 muestras de tamaño N=10, 20, 50 y 100 */
/* con distinta probabilidad 0.01, 0.29, 0.30, 0.40 */

proc iml;
call randseed(-1);
NumTrials = 100;
N = 10;
prob = {0.01,0.29,0.30,0.40};
aa=1;
bb=29;
cc=30;
dd=40;
jac=aa/(aa+bb+cc);
estas=j(1000,3,0);
ta=j(1000,2,0);
err=j(1000,1,0);

do h=1 to 1000;

x=j(N,4,0);
ja=j(N,1,0);

x = RandMultinomial(N,NumTrials,prob);

can=1;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to N;
    a=x[i,1];
    b=x[i,2];
    c=x[i,3];
    d=x[i,4];
    nume=a;
    deno=a+b+c;
    if (deno=0) then goto skip;
    tja=nume/deno;
    ja[can,1] = tja;
    can=can+1;
    skip:
end;
can=can-1;

/* calcula la media y el desvio de la muestra */
suma=0;
suma2=0;
do i = 1 to can;
    suma=suma+ja[i,1];
    suma2=suma2+ja[i,1]*ja[i,1];
end;

estas[h,1]=suma/can;
estas[h,2]=sqrt((suma2-can*estas[h,1]*estas[h,1])/(can-1));
estas[h,3]=can;
end;

/* calcula las medias y los desvios de las 1000 muestras */
med=0;
```

```

var=0;
ser=0;
ene=0;
do i = 1 to 1000;
    err[i,1]=(jac-estas[i,1])*(jac-estas[i,1]);
    ser=ser+err[i,1];
    med=med+estas[i,1];
    var=var+estas[i,2];
    ene=ene+estas[i,3];
end;

prom=med/1000;
des=var/1000;
sero=ser/1000;

print estas err;
print jac prom des ene sero;

zeta=((jac-prom)/(des/sqrt(N)));
print zeta;

/* creates data set PATO */
create pato from err [colname= {ero}];
append from err;
show datasets;
show contents;
close pato;

/* estadística descriptiva */
data pato2;
set pato;
proc univariate;
var ero;
run;

quit;

```

## Programa Propiedad 13: Métrica Prog.758

```
/* prog758Jaccardxy */
/* propiedad métrica de coef. de similitud y disimilitud */
/* calcula el indice de Jaccard (tja) */

proc iml;

can = 1;
nz=10;
coki=j (161700,3,0);
met=j (161700,3,0);
pet=j (161700,3,0);
set=j (161700,3,0);

/* setea los parametros */
/* n=tamaño del vector */
/* p=10 vectores */
/* p1=prob.1 p2=prob.2 ... p10=prob.10 */
n = 100;
p = 10;
z = j (p, n, 0);
tja=j (n, n, 1);
tdja=j (n, n, 1);
tddja=j (n, n, 1);
tdddja=j (n, n, 1);

p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;

/* genera 100 vectores binomiales */
do h=1 to n by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
z[8,g]=z8;
z[9,g]=z9;
```

```

z[10,g]=z10;

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p2);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p2);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g+1]=z1;
z[2,g+1]=z2;
z[3,g+1]=z3;
z[4,g+1]=z4;
z[5,g+1]=z5;
z[6,g+1]=z6;
z[7,g+1]=z7;
z[8,g+1]=z8;
z[9,g+1]=z9;
z[10,g+1]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p3);
z2 = ranbin (seed, 1, p3);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p3);
z5 = ranbin (seed, 1, p3);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p6);
z8 = ranbin (seed, 1, p6);
z9 = ranbin (seed, 1, p6);
z10 = ranbin (seed, 1, p7);
z[1,g+2]=z1;
z[2,g+2]=z2;
z[3,g+2]=z3;
z[4,g+2]=z4;
z[5,g+2]=z5;
z[6,g+2]=z6;
z[7,g+2]=z7;
z[8,g+2]=z8;
z[9,g+2]=z9;
z[10,g+2]=z10;

```

```

seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p10);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p10);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p10);
z10 = ranbin (seed, 1, p9);
z[1,g+3]=z1;
z[2,g+3]=z2;
z[3,g+3]=z3;
z[4,g+3]=z4;

```

```

z[5,g+3]=z5;
z[6,g+3]=z6;
z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;

seed=20;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d (Indice de Jaccard) */
do i = 1 to 100;
  do k = 1 to 100;
    a=0;
    b=0;
    c=0;
    d=0;
    do o = 1 to 10;
      if (z[o,i]<z[o,k]) then c=c+abs(z[o,i]-z[o,k]);
      if (z[o,i]>z[o,k]) then b=b+abs(z[o,i]-z[o,k]);
      a=a+z[o,i]*z[o,k];
      d=d+(1-z[o,i])*(1-z[o,k]);
    end;
    nume=a;
    deno=a+b+c;
    if (deno = 0) then jac=0;
    else jac=nume/deno;
    tja[i,k]=jac;
    tdja[i,k]=1-jac;
    tddja[i,k]=sqrt(1-jac);
    tdddja[i,k]=sqrt(2*(1-jac));
    can=can+1;
  end;
end;

/* hace todas las combinaciones */
coki = allcomb(n, 3);
cant=1;
do g=1 to 161700;
  one=coki[g,1];
  two=coki[g,2];

```

```

tre=coki[g,3];

/* calcula la prop. metrica con tdja */
uno=tdja[one,two];
dos=tdja[one,tre]+tdja[two,tre];
if (uno<=dos) then met[g,1]=0;
else met[g,1]=-1;

uno=tdja[one,tre];
dos=tdja[one,two]+tdja[two,tre];
if (uno<=dos) then met[g,2]=0;
else met[g,2]=-1;

uno=tdja[two,tre];
dos=tdja[two,one]+tdja[one,tre];
if (uno<=dos) then met[g,3]=0;
else met[g,3]=-1;

/* calcula la prop. metrica con tddja */
uno=tddja[one,two];
dos=tddja[one,tre]+tddja[two,tre];
if (uno<=dos) then pet[g,1]=0;
else pet[g,1]=-1;

uno=tddja[one,tre];
dos=tddja[one,two]+tddja[two,tre];
if (uno<=dos) then pet[g,2]=0;
else pet[g,2]=-1;

uno=tddja[two,tre];
dos=tddja[two,one]+tddja[one,tre];
if (uno<=dos) then pet[g,3]=0;
else pet[g,3]=-1;

/* calcula la prop. metrica con tdddja */
uno=tdddja[one,two];
dos=tdddja[one,tre]+tdddja[two,tre];
if (uno<=dos) then set[g,1]=0;
else set[g,1]=-1;

uno=tdddja[one,tre];
dos=tdddja[one,two]+tdddja[two,tre];
if (uno<=dos) then set[g,2]=0;
else set[g,2]=-1;

uno=tdddja[two,tre];
dos=tdddja[two,one]+tdddja[one,tre];
if (uno<=dos) then set[g,3]=0;
else set[g,3]=-1;

cant=cant+1;
end;

/* suma matrices y muestra resultados */
pato=sum(met);
lara=sum(pet);
fran=sum(set);
kant=cant-1;
print kant pato lara fran;

quit;

```

## Programa Propiedad 15: Euclideanidad

### Prog.662

```
/* prog662 indice de Jaccard */
/* propiedad de psd de coef. de sim. y euclideanidad de dis. */
/* para otro indice cambiar la formula de tja */
/* por default calcula psd, para euclideanidad anular calculo sim. */
/* y despintar alguno de los calculos de dis., se pueden usar 3 dis.
*/
```

```
proc iml;
```

```
n = 100;
p = 10;
z = j(p,n,0);
```

```
p1=0.05;
p2=0.10;
p3=0.25;
p4=0.30;
p5=0.45;
p6=0.50;
p7=0.65;
p8=0.70;
p9=0.85;
p10=0.90;
```

```
/* genera 100 vectores binomiales */
```

```
do h=1 to n by 5;
g=h;
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p4);
z5 = ranbin (seed, 1, p5);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p7);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
z[1,g]=z1;
z[2,g]=z2;
z[3,g]=z3;
z[4,g]=z4;
z[5,g]=z5;
z[6,g]=z6;
z[7,g]=z7;
z[8,g]=z8;
z[9,g]=z9;
z[10,g]=z10;
```

```
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p2);
z5 = ranbin (seed, 1, p1);
z6 = ranbin (seed, 1, p2);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p8);
z9 = ranbin (seed, 1, p9);
z10 = ranbin (seed, 1, p10);
```

```
z[1,g+1]=z1;
z[2,g+1]=z2;
z[3,g+1]=z3;
z[4,g+1]=z4;
z[5,g+1]=z5;
z[6,g+1]=z6;
z[7,g+1]=z7;
z[8,g+1]=z8;
z[9,g+1]=z9;
z[10,g+1]=z10;
```

```
seed=h;
z1 = ranbin (seed, 1, p3);
z2 = ranbin (seed, 1, p3);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p3);
z5 = ranbin (seed, 1, p3);
z6 = ranbin (seed, 1, p6);
z7 = ranbin (seed, 1, p6);
z8 = ranbin (seed, 1, p6);
z9 = ranbin (seed, 1, p6);
z10 = ranbin (seed, 1, p7);
z[1,g+2]=z1;
z[2,g+2]=z2;
z[3,g+2]=z3;
z[4,g+2]=z4;
z[5,g+2]=z5;
z[6,g+2]=z6;
z[7,g+2]=z7;
z[8,g+2]=z8;
z[9,g+2]=z9;
z[10,g+2]=z10;
```

```
seed=h;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p2);
z3 = ranbin (seed, 1, p3);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p10);
z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p10);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p10);
z10 = ranbin (seed, 1, p9);
z[1,g+3]=z1;
z[2,g+3]=z2;
z[3,g+3]=z3;
z[4,g+3]=z4;
z[5,g+3]=z5;
z[6,g+3]=z6;
z[7,g+3]=z7;
z[8,g+3]=z8;
z[9,g+3]=z9;
z[10,g+3]=z10;
```

```
seed=20;
z1 = ranbin (seed, 1, p1);
z2 = ranbin (seed, 1, p9);
z3 = ranbin (seed, 1, p1);
z4 = ranbin (seed, 1, p9);
z5 = ranbin (seed, 1, p1);
```

```

z6 = ranbin (seed, 1, p9);
z7 = ranbin (seed, 1, p1);
z8 = ranbin (seed, 1, p9);
z9 = ranbin (seed, 1, p1);
z10 = ranbin (seed, 1, p9);
z[1,g+4]=z1;
z[2,g+4]=z2;
z[3,g+4]=z3;
z[4,g+4]=z4;
z[5,g+4]=z5;
z[6,g+4]=z6;
z[7,g+4]=z7;
z[8,g+4]=z8;
z[9,g+4]=z9;
z[10,g+4]=z10;
end;

/* calcula a, b, c, d */
/* Indice de similitud de Jaccard (tja) */
tja=j(n,n,1);
do k = 1 to 100;
    do l = 1 to 100;
        a=0;
        b=0;
        c=0;
        d=0;
        do i = 1 to 10;
            if (z[i,k]<z[i,l]) then c=c+abs(z[i,k]-z[i,l]);
            if (z[i,k]>z[i,l]) then b=b+abs(z[i,k]-z[i,l]);
            a=a+z[i,k]*z[i,l];
            d=d+(1-z[i,k])*(1-z[i,l]);
        end;

        nume=a;
        deno=a+b+c;
        if (deno = 0) then jac=0;
        else jac=nume/deno;
        tja[k,l]=jac;
    end;
end;

/* calcula la disimilitud 1 (tdja=1-tja) */
/* calcula la disimilitud 2 (tddja=sqrt(tdja)) */
/* calcula la disimilitud 3 (tdddja=sqrt(2*tdja)) */
tdja=1-tja;
tddja=sqrt(tdja);
tdddja=sqrt(2*tdja);

/* imprime todo *
print tja tdja tddja tdddja; */

/* calcula la matriz A para similitud */
ata=tja;

/* calcula la matriz A para disimilitud 1 *
ata = -0.5*((tdja)##2);

/* calcula la matriz A para disimilitud 2 *
ata = -0.5*((tddja)##2);

/* calcula la matriz A para disimilitud 3 *

```

```

ata = -0.5*((tdddja)##2);

/* calcula las medias de filas */
fa=j(n,1,0);
do i = 1 to 100;
    rf=0;
    do k = 1 to 100;
        rf=rf+ata[i,k];
    end;
fa[i,1]=rf;
end;
fha=fa/n;

/* calcula las medias de columnas */
ca=j(1,n,0);
do k = 1 to 100;
    rc=0;
    do i = 1 to 100;
        rc=rc+ata[i,k];
    end;
ca[1,k]=rc;
end;
cha=ca/n;

/* calcula la media total */
tt=0;
do i = 1 to 100;
    do k = 1 to 100;
        tt=tt+ata[i,k];
    end;
end;
tht=tt/(n*n);

/* calcula la matriz DELTA */
de=j(n,n,0);
do i = 1 to 100;
    do k = 1 to 100;
        de[i,k]=(ata[i,k]-fha[i,1]-cha[1,k]+tht);
    end;
end;

/* calcula el determinante y la traza de DELTA */
dta=det(de);
tra=trace(de);

/* imprime todo */
print fha cha tht de dta tra; */

/* hace el analisis eigen de DELTA */
call eigen(a,b,de);
sua=sum(a);
print a;
/*print b de sua*/

/* calcula las coord. ppales. */
at=j(1,n,1);
ats=j(1,n,1);
cc=j(n,n,1);
bt=j(n,n,1);
at=t(a);
k=1;

```

```

do while (k<=n);
if (at[k]<0) then at[k]=0;
k=k+1;
end;
ats=sqrt(at);

bt=t(b);
i=1;
do while (i<=n);
j=1;
do while (j<=n);
cc[i,j]=bt[i,j]*ats[i];
j=j+1;
end;
i=i+1;
end;

/* imprime todo *
print at ats bt cc; */

quit;

```

## Programa Ejemplo: Datos reales

### Prog.678

```
/* prog678Jaccard Transecta de datos reales */  
/* calcula el indice de Jaccard (tja) */  
/* para otro indice cambiar la formula de tja */  
/* se pueden usar 3 disimilitudes para PCO */
```

```
proc iml;
```

```
n = 27;
```

```
p = 65;
```

```
z={1 1 1 1 1 1 1 1 1 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 0 0 1 0 0 0 0 1 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 0 1 0 1 0 0 0 1 1 1  
1 1 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 0 1 0 1 0 0 0 0 1 0  
0 0 1 1 1 1 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 0 1 0 1 0 0 0 0 1 0  
0 0 1 1 1 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 1 0 1 0 0 0 0 0 0 0  
0 0 1 1 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0, 0 0  
1 1 0 1 0 1 0 0 0 0 0 0  
0 0 1 1 1 0 0 0 0 1 1 0  
0 0 0 0 0 0 0 0 0 0 0 0
```

	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	1	0	1	0	1	0	1
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	1	0	0	0
	0	0	1	0	0	1	1	0	1	0	1
	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	1	0	0	0
	0	0	0	0	0	1	0	0	1	0	1
	0	1	1	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	1	1	0	0	0	0
	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	1	0	0	1	0	0	0
	0	0	0	0	0	1	1	1	1	0	0
	0	0	0	1	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0
	0	1	1	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	1	0	1	0	0	0
	0	0	0	0	0	1	0	0	1	0	0
	0	1	0	1	1	1	0	0	1	0	0
	0	0	0	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	1	0	1	0	0	0
	0	0	0	0	0	1	0	0	1	0	0
	0	1	1	1	0	1	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	0
	0	1	0	1	1	1	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0
1	1	0	1	0	1	0	0	0	1	0	0
	1	0	0	0	0	0	0	0	1	0	0
	1	0	0	0	1	1	0	0	0	1	1
	0	1	1	0	0	0	0	0	0	0	0

```

0      0      0      0      0      0      0      0      0      1      0      0
0      0      0      0      0      0      0      0      0      0,    0      0
1      1      0      0      1      1      0      0      0      0      0      0
1      1      0      0      0      0      0      1      1      1      0      0
1      1      1      1      1      1      1      1      1      0      0      1
0      0      1      1      1      1      1      1      1      0      0      1
0      0      0      0      0      0      0      0      0      0      0      0
1      1      1      0      0      0      0      0      0      0,    0      0
1      1      0      0      0      1      0      0      0      1      0      0
1      1      0      0      0      0      0      1      0      1      0      0
0      0      1      0      1      1      1      1      0      0      1      1
0      1      1      0      0      1      0      0      0      0      1      1
1      1      0      0      1      1      0      0      0      0,    0      0
1      1      0      1      0      0      0      0      0      1      0      0
1      0      1      0      0      1      1      0      0      0      0      0
0      0      1      0      1      1      1      0      0      0      1      1
0      0      1      0      1      0      0      1      1      0      1      1
0      0      0      0      0      0      0      0      0      0,    0      0
0      0      1      0      0      0      0      0      0      1      0      0
0      0      0      0      0      0      0      0      0      0      0      0
0      0      0      0      1      1      0      0      0      0      1      1
0      0      0      0      0      0      0      0      0      0,    0      0
0      0      0      0      0      0      0      0      0      1      0      0
0      0      0      0      0      0      0      0      0      0      0      0
0      0      0      0      0      1      0      0      0      0      1      1
0      0      0      0      0      0      0      0      0      0,    0      0
0      0      1      1      0      1      0      1      0      0      0      0
0      0      0      0      0      0      0      0      0      0,    0      0
0      1      0      0      0      0      0      0      0      0      0      0
0      0      0      0      0      0      0      0      0      0      1      1
0      0      0      0      1      0      0      0      0      0      0      0
0      1      1      0      0      0      0      0      0      0      0      0
0      0      0      0      0      1      1      1      1      1,    0      0

```

```

/* calcula a, b, c, d */
/* Indice de similitud de Jaccard (tja) */
tja=j(n,n,1);
k=1;
do while (k<=n);
  l=1;
  do while (l<=n);
    a=0;
    b=0;
    c=0;

```

```

        d=0;
        i=1;
        do while (i<=p);
        if (z[k,i]<z[l,i]) then c=c+abs(z[k,i]-z[l,i]);
        if (z[k,i]>z[l,i]) then b=b+abs(z[k,i]-z[l,i]);
        a=a+z[k,i]*z[l,i];
        d=d+(1-z[k,i])*(1-z[l,i]);
        i=i+1;
        end;
        tja[k,l]=a/(a+b+c);
        l=l+1;
    end;
k=k+1;
end;

/* calcula las disimilitudes */
tdja=1-tja;
tddja=sqrt(tdja);
tdddja=sqrt(2*tdja);

/* calcula la matriz A para similitud *
ata=tja;
print ata;

/* calcula la matriz A para disimilitud 1 tdja=1-tja *
ata = -0.5*((tdja)##2);
print ata;

/* calcula la matriz A para disimilitud 2 tddja=sqrt(tdja)*/
ata = -0.5*((tddja)##2);
print ata;

/* calcula la matriz A para disimilitud 3 tdddja=sqrt(2*tdja)*
ata = -0.5*((tdddja)##2);
print ata;

/* calcula las medias de filas */
fa=j(n,1,0);
i=1;
do while (i<=n);
    rf=0;
    k=1;
    do while (k<=n);
        rf=rf+ata[i,k];
        k=k+1;
    end;
    fa[i,1]=rf;
    i=i+1;
end;
fha=fa/n;

/* calcula las medias de columnas */
ca=j(1,n,0);
k=1;
do while (k<=n);
    rc=0;
    i=1;
    do while (i<=n);
        rc=rc+ata[i,k];
        i=i+1;
    end;
end;

```

```

ca[1,k]=rc;
k=k+1;
end;
cha=ca/n;

/* calcula la media total */
tt=0;
i=1;
do while (i<=n);
    k=1;
    do while (k<=n);
        tt=tt+ata[i,k];
        k=k+1;
    end;
    i=i+1;
end;
tht=tt/(n*n);

/* calcula la matriz DELTA */
de=j(n,n,0);
i=1;
do while (i<=n);
    k=1;
    do while (k<=n);
        de[i,k]=(ata[i,k]-fha[i,1]-cha[1,k]+tht);
        k=k+1;
    end;
    i=i+1;
end;

/* calcula el determinante y la traza de DELTA */
dta=det(de);
tra=trace(de);

/* hace el analisis eigen de DELTA */
call eigen(a,b,de);
sua=sum(a);
print a, b, sua;

/* calcula las coord. ppales. */
at=j(1,n,1);
ats=j(1,n,1);
cc=j(n,n,1);
bt=j(n,n,1);
at=t(a);
k=1;
    do while (k<=n);
        if (at[k]<0) then at[k]=0;
        k=k+1;
    end;
ats=sqrt(at);

bt=t(b);
i=1;
do while (i<=n);
    j=1;
    do while (j<=n);
        cc[i,j]=bt[i,j]#ats[i];
        j=j+1;
    end;
    i=i+1;
end;

```

```

end;

/* imprime todo */
print at ats bt cc;

```

### Programa Ejemplo: Coenocline Prog.772

```

/* prog772Jaccard Coenocline */
/* calcula el indice de Jaccard (tja) */
/* para otro indice cambiar la formula de tja */
/* se pueden usar 3 disimilitudes para PCO */

```

```

proc iml;
n = 30;
p = 60;
z={0 1 0 1 0 0 0 0 1 1 0 0
    0 1 0 0 1 0 0 0 0 1 0 0 1
    1 1 1 0 0 1 1 1 1 1 1 1 1
    0 1 1 0 0 1 1 0 0 0 0 0 1
    0 1 1 1 1 1 1 0 1 0 0 0 0
    1 0 0 1,
0 1 0 1 0 0 0 0 1 0 0 0 0
    0 1 0 0 1 0 1 0 1 0 1 0 1
    1 1 1 0 0 0 1 1 1 1 0 1
    0 1 1 0 0 1 1 1 0 0 0 0 0
    0 1 1 1 1 1 1 0 1 0 0 0 0
    1 0 0 1,
0 1 0 1 0 0 0 1 0 1 0 0 0
    0 1 0 1 1 0 0 0 1 0 0 0 1
    0 0 1 0 0 0 1 1 0 0 0 0 0
    0 1 1 0 1 0 0 0 0 0 0 0 0
    1 0 0 0,
0 0 1 0 1 0 0 0 0 0 1 0 0
    0 0 1 0 0 0 0 0 1 0 1 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 1 0 0 0
    0 0 0 0,
1 0 0 0 0 1 1 1 0 1 0 0 0
    1 0 0 1 0 1 0 0 0 0 0 0 0
    0 0 0 1 1 0 0 0 1 0 1 0 0
    1 1 0 1 1 1 0 0 0 1 1 0 0
    1 0 0 1 0 0 0 1 1 1 0 1 1
    0 1 1 1,
0 0 1 1 0 0 0 0 1 0 1 1 1
    0 1 0 0 0 0 0 0 1 1 0 0 0
    0 1 0 0 0 0 0 0 0 0 0 0 0
    0 0 1 0 0 1 1 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 1 0 0 0
    1 0 0 0,
1 1 0 0 0 0 1 1 0 1 0 0 0
    1 0 0 1 0 1 0 0 0 0 0 0 0

```

	0	0	0	1	1	0	0	1	1	1	0
	1	1	0	1	0	0	0	1	1	1	0
	1	1	0	1	0	0	1	1	0	1	1
1	0	1	1	1,							
	0	0	0	0	0	1	1	0	1	0	0
	1	0	0	1	0	1	0	0	0	0	0
	0	0	0	1	1	0	0	1	0	1	0
	1	0	0	1	0	0	0	0	1	1	0
	1	0	0	1	0	0	1	1	0	1	1
0	0	1	1	0,							
	0	1	0	1	0	0	0	1	0	1	1
	0	0	0	0	0	0	0	1	1	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0
0	0	1	1	1	0	0	0	1	0	1	1
	0	0	0	0	0	0	0	1	1	0	0
	0	1	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	1	1	0	0	0	0
	0	0	0	0	1	0	0	0	1	0	0
0	1	0	0	0,							
	0	0	1	1	0	0	0	1	0	1	0
	0	1	0	0	1	0	0	0	1	0	1
	1	1	1	0	0	0	0	1	1	0	1
	0	0	0	0	0	1	1	0	0	0	1
	0	1	1	1	1	0	0	0	0	0	0
0	1	0	0	0,							
	0	0	0	1	0	0	0	0	0	1	0
	0	0	1	0	0	0	0	1	1	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	1	0	1	0	0
	1	0	0	1	0	1	0	0	0	0	0
	0	0	0	1	1	0	0	1	0	1	0
	1	1	0	1	1	0	0	1	1	1	0
	1	0	0	1	0	0	1	1	0	0	1
0	0	1	1	0,							
	0	0	1	0	1	0	0	1	0	1	1
	0	0	0	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0,							
	0	0	0	0	0	1	1	0	1	0	0
	1	0	0	1	0	1	0	0	0	0	0
	0	0	0	1	1	0	0	1	1	1	0
	1	1	0	1	1	0	0	1	1	1	0
	1	1	0	1	0	0	1	1	0	0	1
0	0	1	1	0,							
	0	1	0	1	0	0	0	0	0	1	0
	0	0	1	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0,							
	0	0	1	1	0	0	0	0	0	1	0
	0	0	1	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0

	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0,							
	1	0	0	0	1	1	1	0	1	0	0
	1	0	0	1	0	1	0	0	0	0	1
	0	0	0	1	1	0	0	1	1	1	0
	1	1	0	1	0	0	0	0	1	1	1
	1	0	0	1	0	0	0	1	0	1	1
0	0	1	1	1,							
	0	1	0	1	0	0	0	0	0	1	0
	0	0	1	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0,							
1	0	0	0	0	1	1	1	0	1	0	0
	1	0	0	1	0	1	0	0	0	0	0
	0	0	0	1	1	0	0	1	0	1	0
	1	1	0	1	1	0	0	0	1	1	0
	1	0	0	1,	0	0	1	1	0	1	1
0	0	1	1	1,							
	1	0	1	0	1	0	1	0	1	0	0
	0	1	0	1	1	0	0	0	0	0	1
	1	1	1	1	1	0	1	1	1	1	0
	0	1	1	0	0	1	1	0	0	0	1
	0	1	1	1	1	1	0	1	0	0	1
0	1	1	0	1,							
	0	1	0	1	0	0	0	0	0	1	0
	0	0	1	0	0	0	0	1	1	1	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0,							
0	1	0	1	0	1	0	0	0	1	0	0
	0	1	0	1	1	0	1	0	0	0	1
	1	1	0	1	1	1	1	1	1	1	0
	0	1	1	0	0	1	0	0	1	1	1
	0	1	1	1	1	0	0	1	0	0	1
0	1	1	0	1,							
	0	1	0	1	0	0	0	1	0	1	1
	0	0	0	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0,							
0	0	1	0	1	0	0	0	1	0	1	1
	0	0	0	0	0	0	0	1	1	0	0
	0	1	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	0	0
	0	0	1	0	0	1	1	0	0	0	0

```

0      0      0      0      0      0      0      0      0      1      0      0
      1      0      0      0,
0      0      1      0      1      0      0      0      0      0      1      0
      0      0      1      0      0      0      0      0      1      1      1      0
      0      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      0,
0      0      1      0      1      0      0      0      1      0      1      1
      0      1      0      0      0      0      0      0      1      1      0      0
      0      1      0      0      0      0      0      0      0      0      0      0
      0      0      0      0      0      1      1      0      0      0      0      0
      0      0      0      0      0      0      0      0      0      1      0      0
      1      0      0      0};

```

```

/* calcula a, b, c, d */
/* Indice de similitud de Jaccard (tja) */
tja=j(n,n,1);
k=1;
do while (k<=n);
  l=1;
  do while (l<=n);
    a=0;
    b=0;
    c=0;
    d=0;
    i=1;
    do while (i<=p);
      if (z[k,i]<z[l,i]) then c=c+abs(z[k,i]-z[l,i]);
      if (z[k,i]>z[l,i]) then b=b+abs(z[k,i]-z[l,i]);
      a=a+z[k,i]*z[l,i];
      d=d+(1-z[k,i])*(1-z[l,i]);
      i=i+1;
    end;
    tja[k,l]=a/(a+b+c);
    l=l+1;
  end;
  k=k+1;
end;

/* calcula las disimilitudes */
tdja=1-tja;
tddja=sqrt(tdja);
tdddja=sqrt(2*tdja);

/* calcula la matriz A para similitud */
ata=tja;
print ata;

/* calcula la matriz A para disimilitud 1 tdja=1-tja */
ata = -0.5*((tdja)##2);
print ata;

/* calcula la matriz A para disimilitud 2 tddja=sqrt(tdja) */
ata = -0.5*((tddja)##2);
print ata;

/* calcula la matriz A para disimilitud 3 tdddja=sqrt(2*tdja) */
ata = -0.5*((tdddja)##2);
print ata;

```

```

/* calcula las medias de filas */
fa=j(n,1,0);
i=1;
do while (i<=n);
    rf=0;
    k=1;
    do while (k<=n);
        rf=rf+ata[i,k];
        k=k+1;
    end;
    fa[i,1]=rf;
    i=i+1;
end;
fha=fa/n;

/* calcula las medias de columnas */
ca=j(1,n,0);
k=1;
do while (k<=n);
    rc=0;
    i=1;
    do while (i<=n);
        rc=rc+ata[i,k];
        i=i+1;
    end;
    ca[1,k]=rc;
    k=k+1;
end;
cha=ca/n;

/* calcula la media total */
tt=0;
i=1;
do while (i<=n);
    k=1;
    do while (k<=n);
        tt=tt+ata[i,k];
        k=k+1;
    end;
    i=i+1;
end;
tht=tt/(n*n);

/* calcula la matriz DELTA */
de=j(n,n,0);
i=1;
do while (i<=n);
    k=1;
    do while (k<=n);
        de[i,k]=(ata[i,k]-fha[i,1]-cha[1,k]+tht);
        k=k+1;
    end;
    i=i+1;
end;

/* calcula el determinante y la traza de DELTA */
dta=det(de);
tra=trace(de);

/* hace el analisis eigen de DELTA */
call eigen(a,b,de);

```

```

sua=sum(a);
print a, b, sua;

/* calcula las coord. ppales. */
at=j(1,n,1);
ats=j(1,n,1);
cc=j(n,n,1);
bt=j(n,n,1);
at=t(a);
k=1;
  do while (k<=n);
    if (at[k]<0) then at[k]=0;
    k=k+1;
  end;
ats=sqrt(at);

bt=t(b);
i=1;
do while (i<=n);
  j=1;
  do while (j<=n);
    cc[i,j]=bt[i,j]#ats[i];
    j=j+1;
  end;
  i=i+1;
end;

/* imprime todo */
print at ats bt cc;

```

## Programa Ejemplo: Coenoplano Prog.882

```

/* prog882Jaccard Coenoplano*/
/* calcula el indice de Jaccard (tja) */
/* para otro indice cambiar la formula de tja */
/* se pueden usar 3 disimilitudes para PCO */

```

```

proc iml;
n = 30;
p = 59;
z={1  0    1    0    0    0    0    0    0    0    0    0
    0    0    1    0    0    0    0    0    0    0    1    1
    0    0    0    1    1    0    0    1    1    0    0    1
    0    0    0    0    0    0    1    1    0    0    0    0
    1    1    0    0    0    0    0    0    0    0    0    0
    0    0    1,
0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    1    0    0    1
    0    1    0    0    1    0    0    0    0    0    0    0
    0    0    0    0    0    1    0    1    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0,
1    0    0    0    0    0    0    0    1    0    0    1
    0    0    0    0    1    0    0    0    0    1    1    0
    0    0    0    0    0    0    0    0    0    0    1    0
    0    1    0    0    0    0    0    0    0    0    0    1
    0    0    0    0    0    1    1    0    0    0    0    0
    0    0    0,

```

0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	1	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	1	0	0	1	0
	0	0	0	0	0	0	1	1	1	1	1	0
	1	1	0	0	0	0	0	0	0	0	0	0
	0	0	1,									
0	0	0	1	1	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	1	1	0	0	0	0
	1	0	0	0	0	0	1	0	0	0	0	0
	0	0	0	1	1	0	0	0	0	1	0	0
	0	0	0	1	0	0	0	0	0	0	0	1
	0	1	0,									
0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	1	0	0	1	0
	0	1	0	0	1	0	0	0	1	0	1	0
	0	0	0	0	0	1	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	1,									
1	0	0	0	1	1	1	0	0	1	1	1	1
	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	1	0	0
	1	0	0	0	0	0	0	0	1	0	1	0
	0	0	0	0	0	1	0	0	0	0	0	0
	1	0	0,									
0	0	0	0	1	0	1	0	0	0	0	0	0
	0	1	0	0	0	0	1	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	1	0	1	1	1	1	0
	0	0	0	1	1	0	0	0	0	0	0	0
	0	0	1,									
0	0	0	0	0	0	0	0	1	0	0	1	0
	0	0	0	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0,									
0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	1	0	0	1	0
	0	0	0	0	1	0	0	0	1	0	1	0
	0	0	0	0	0	1	1	1	0	0	0	0
	1	0	0	0	1	0	0	0	0	0	0	0
	0	0	1,									
0	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	1	1	1	0	1	0	0
	6	1	0	0	1	0	0	0	0	0	0	0
	0	0	1,									
0	0	0	0	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	1	0	0
	0	1	0	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0,									
1	0	0	0	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	1	0	0
	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	1	0	0	0	0	0
	0	0	0,									

0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	1	0	1	0
	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	1,	0	0	0	0	0	0	0	0	0
	0	0	0	1	1	0	0	0	0	0	0	1
	0	0	0	0	0	1	1	0	1	1	0	1
	0	0	1	0	0	0	1	1	1	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
	0	0	1,	0	0	0	0	1	0	1	1	1
0	0	0	0	0	0	0	0	0	1	0	1	0
	0	0	0	0	0	0	0	0	0	1	0	0
	0	0	1	0	0	0	0	0	0	0	1	0
	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0,	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	1	0	0	0	1
	0	1	0	0	1	0	0	0	1	0	0	1
	0	0	0	0	0	0	1	0	0	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1,	0	0	0	0	0	1	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	1	0	1	1	1	0
	0	1	0	1	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	1	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	1	0
	0	0	0	0	1	0	0	0	0	0	1	0
	0	0	0,	0	1	0	0	0	0	0	0	1

```

0      0      0      0      0      0      0      1      0      1      0      0
      1      1      0      0      0      0      0      0      0      0      0      0
      1      0      0      0      0      0      0      1      0      0      0      0
      1      0      0      0      1      0      0      0      0      0      0      0
      0      0      1      0      0      0      0      0      1      0      1      0
      0      1      0,
0      0      0      1      1      0      1      0      0      0      0      0
      0      0      0      0      0      0      0      1      1      0      0      0
      1      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      1      1      0      0      0      0      1      1      0
      0      0      0      1      1      0      0      0      0      0      0      1
      0      1      0,
1      0      0      0      0      1      1      0      0      1      0      1
      0      1      0      0      0      0      0      0      0      0      1      0
      0      0      0      0      0      0      0      0      0      0      0      0
      1      0      0      0      0      0      0      0      0      1      0      1
      0      0      0      0      1      1      0      0      0      0      0      0
      1      0      0,
0      1      0      0      0      0      0      0      0      0      0      0
      1      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      0      1      0      0      0      0      0      0      0
      0      0      1      0      0      0      0      0      1      0      1      0
      0      0      0,
0      0      0      0      1      0      0      0      0      1      0      0
      1      1      0      0      0      0      0      1      0      0      0      0
      1      0      0      0      0      0      0      1      0      0      0      0
      1      0      0      0      1      0      0      0      0      0      0      0
      0      0      1      0      0      0      0      0      1      0      1      0
      0      1      0,
0      0      0      0      0      0      0      0      0      0      0      0
      0      0      1      0      1      0      0      0      0      1      1      0
      0      0      0      0      0      0      0      0      1      0      0      0
      0      0      1      0      0      0      0      1      0      0      0      0
      0      1      0      0      0      0      0      1      0      0      0      0
      0      0      0,
0      0      0      0      0      0      0      0      0      0      0      0
      0      0      0      0      0      0      0      0      1      0      0      1
      0      1      0      0      1      0      0      0      0      0      0      0
      0      0      0      0      0      0      1      0      0      0      0      0
      0      0      0      0      0      0      0      0      0      0      0      0
      0      0      0};

```

```

/* calcula a, b, c, d */
/* Indice de similitud de Jaccard (tja) */
tja=j(n,n,1);
k=1;
do while (k<=n);
  l=1;
  do while (l<=n);
    a=0;
    b=0;
    c=0;
    d=0;
    i=1;
    do while (i<=p);
      if (z[k,i]<z[l,i]) then c=c+abs(z[k,i]-z[l,i]);
      if (z[k,i]>z[l,i]) then b=b+abs(z[k,i]-z[l,i]);
      a=a+z[k,i]*z[l,i];
      d=d+(1-z[k,i])*(1-z[l,i]);
      i=i+1;
    end do;
    l=l+1;
  end do;
  k=k+1;
end do;

```

```

        end;
        tja[k,1]=a/(a+b+c);
        l=l+1;
    end;
k=k+1;
end;

/* calcula las disimilitudes */
tdja=1-tja;
tddja=sqrt(tdja);
tdddja=sqrt(2*tdja);

/* calcula la matriz A para similitud *
ata=tja;
print ata;

/* calcula la matriz A para disimilitud 1 tdja=1-tja *
ata = -0.5*((tdja)##2);
print ata;

/* calcula la matriz A para disimilitud 2 tddja=sqrt(tdja) */
ata = -0.5*((tddja)##2);
print ata;

/* calcula la matriz A para disimilitud 3 tdddja=sqrt(2*tdja) *
ata = -0.5*((tdddja)##2);
print ata;

/* calcula las medias de filas */
fa=j(n,1,0);
i=1;
do while (i<=n);
    rf=0;
    k=1;
    do while (k<=n);
        rf=rf+ata[i,k];
        k=k+1;
    end;
fa[i,1]=rf;
i=i+1;
end;
fha=fa/n;

/* calcula las medias de columnas */
ca=j(1,n,0);
k=1;
do while (k<=n);
    rc=0;
    i=1;
    do while (i<=n);
        rc=rc+ata[i,k];
        i=i+1;
    end;
ca[1,k]=rc;
k=k+1;
end;
cha=ca/n;

/* calcula la media total */
tt=0;
i=1;

```

```

do while (i<=n);
    k=1;
    do while (k<=n);
        tt=tt+ata[i,k];
        k=k+1;
    end;
i=i+1;
end;
tht=tt/(n*n);

/* calcula la matriz DELTA */
de=j(n,n,0);
i=1;
do while (i<=n);
    k=1;
    do while (k<=n);
        de[i,k]=(ata[i,k]-fha[i,1]-cha[1,k]+tht);
        k=k+1;
    end;
i=i+1;
end;

/* calcula el determinante y la traza de DELTA */
dta=det(de);
tra=trace(de);

/* hace el analisis eigen de DELTA */
call eigen(a,b,de);
sua=sum(a);
print a, b, sua;

/* calcula las coord. ppales. */
at=j(1,n,1);
ats=j(1,n,1);
cc=j(n,n,1);
bt=j(n,n,1);
at=t(a);
k=1;
    do while (k<=n);
        if (at[k]<0) then at[k]=0;
        k=k+1;
    end;
ats=sqrt(at);

bt=t(b);
i=1;
do while (i<=n);
    j=1;
    do while (j<=n);
        cc[i,j]=bt[i,j]#ats[i];
        j=j+1;
    end;
i=i+1;
end;

/* imprime todo */
print at ats bt cc;

```