

EL ERROR DE OBSERVACION Y SU INFLUENCIA EN LOS ANALISIS MORFOLOGICOS DE RESTOS OSEOS HUMANOS. DATOS DE VARIACION DISCRETA

*Paula N. González¹
Valeria Bernal¹
Sergio I. Perez^{1,2}
Mariano Del Papa¹
Florencia Gordon¹
Gabriela Ghidini¹*

PALABRAS CLAVE: Rasgos no métricos, Esqueleto craneofacial, Error intra e interobservador, Relaciones poblacionales

RESUMEN: A partir de la década de 1960 se incrementó el empleo de los rasgos no métricos del cráneo en los análisis de relaciones poblacionales. Si bien uno de los supuestos en que se basó el uso de estos rasgos es la facilidad de estandarizar su registro, varios trabajos sugieren que el error interobservador en los rasgos discretos es elevado. En consecuencia, los objetivos del trabajo son evaluar el error intra e interobservador en el registro de rasgos craneales no métricos y analizar el efecto del error interobservador sobre las distancias biológicas calculadas. Con este fin, se analizó una muestra arqueológica procedente del Valle inferior del Río Negro -10 con deformación pseudocircular y 10 con deformación

1 Departamento Científico de Antropología del Museo de La Plata. Facultad de Ciencias Naturales y Museo. Universidad Nacional de La Plata. Paseo del Bosque s/n. 1900 La Plata. Argentina.
e-mail: pgonzalez@museo.fcnym.unlp.edu.ar

2 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

planolámbdica-. Se seleccionaron ocho variables discretas del cráneo que fueron relevadas por cuatro observadores. El error de observación fue evaluado mediante un diseño experimental de bloques aleatorios con tres repeticiones y el empleo del índice Kappa y la prueba de McNemar. La distancia biológica entre las dos muestras se estimó mediante la Medida Media de Divergencia y un análisis de escalamiento multidimensional. Los resultados obtenidos indican que el error intraobservador disminuyó mediante la aplicación del diseño experimental, que el error interobservador se incrementó en las sucesivas series y que las diferencias entre los observadores alteraron los resultados de las medidas de distancia calculadas entre las muestras. *Rev. Arg. Antrop. Biol.* 6(1): 35-46, 2004.

KEY WORDS: Nonmetric traits, Cranial-facial skeleton, Intra and inter-observer error, Population relationships

ABSTRACT: Since the early 1960s, there was an increasing interest in the application of non-metric cranial traits to the analysis of relationships between populations. One of the assumptions for the use of these traits is based upon the simplicity to standardize the recordings. However, several papers suggest that the interobserver error on such recordings is high. Therefore, the goals of this paper are to evaluate the intra and interobserver error on the scoring of non-metric cranial traits, as well as to analyze the effect of the interobserver error in the biological distances estimated with them. An archeological sample (n=20) from the lower stretch Valley of Río Negro Valley (Río Negro Province, Argentina) was analyzed. Eight discrete variables from the skull were recorded independently by four observers. The observation error was evaluated by means of a randomized complete blocks design, with three repetitions, and employing the Kappa index and the McNemar test. The biological distance between the two samples was estimated through the Mean Measure of Divergence and a multidimensional scaling test. Our results indicate that 1) the intraobserver error diminishes with the application of observational designs; 2) the interobserver error increases in successive series and; 3) the differences among observers modify the results of calculated biodistance between samples. *Rev. Arg. Antrop. Biol.* 6(1): 35-46, 2004.

INTRODUCCION

El empleo de rasgos no métricos del cráneo en los estudios de relaciones intra e interpoblacionales se incrementó a partir de la década de 1960 (Berry y Berry, 1967; Corrucini, 1974; Ossenberg, 1976; Sciulli, 1990; Pardoe, 1991; Konigsberg et al., 1993; Hanihara et al., 2003; entre otros). Si bien dentro de la

antropología estos rasgos comenzaron a utilizarse a principios del siglo XX (Outes, 1911; Hooton, 1918), el tratamiento de los datos fue básicamente descriptivo hasta la incorporación de medidas de distancia poblacional basadas en el análisis de la frecuencia de rasgos no métricos (v.g. Medida Media de Divergencia). La amplia utilización de los caracteres discretos se basó en el supuesto de que hay un fuerte componente genético en su expresión fenotípica y en la facilidad para estandarizar el registro de los mismos (Berry y Berry, 1967). Sin embargo, diversos trabajos sugieren que, a pesar de sus bondades, los rasgos no métricos no están exentos de un alto grado de error interobservador, incluso mayor que en los datos métricos (Zegura, 1975; Molto, 1979; Saunders, 1989; Gualdi-Russo et al., 1999). Por lo tanto, aún cuando se emplean técnicas estandarizadas para la recolección de este tipo de información el uso de datos obtenidos por diferentes investigadores puede alterar significativamente las comparaciones poblacionales realizadas (Finnegan y Rubison, 1980). En este sentido, Saunders (1989) plantea que el bajo poder discriminante de los rasgos no métricos en los estudios de poblaciones esqueléticas puede atribuirse en muchos casos a las diferentes variables utilizadas y a las dificultades metodológicas para su descripción y registro. Considerando que el empleo de bases de datos generadas por distintos investigadores es una práctica extendida dentro de los análisis de relaciones biológicas (González-José et al., 2001; Hanihara e Ishida, 2001a; b; entre otros), es necesario elaborar estrategias orientadas a evaluar e incrementar la precisión entre las observaciones cuando han sido realizadas por uno o más observadores. En consecuencia, los objetivos del presente trabajo son: 1.- evaluar el error intra e interobservador en el registro de rasgos craneales no métricos y la influencia que tiene el entrenamiento de los observadores sobre cada tipo de error; 2.- analizar el efecto que tiene el error interobservador en el análisis de relaciones biológicas entre muestras arqueológicas. Con este fin se analizaron dos muestras procedentes del Valle inferior del Río Negro para las cuales hay evidencia suficiente (tipo de deformación craneana, fechados radiocarbónicos, tipos de entierros, entre otras) que permite sostener que ambas proceden de distintos momentos del Holoceno (Berón y Baffi, 2003; Barrientos y Perez, 2004). Las principales hipótesis a evaluar son: A) no hay diferencias entre los observadores y por lo tanto las únicas diferencias encontradas corresponden a las existentes entre las dos muestras analizadas y B) las diferencias encontradas corresponden a diferencias debidas al entrenamiento de los observadores, es decir, no hay diferencias entre las muestras relevadas por los observadores entrenados.

MATERIAL Y METODOS

La muestra analizada se seleccionó aleatoriamente de la Colección del Departamento Científico de Antropología del Museo de La Plata (FCNyM. UNLP) mediante el empleo del programa Random Number Generator Pro V1.26 (Segobit Software, 2001). El conjunto así seleccionado consistió en 20 cráneos procedentes del Valle del Río Negro -10 con deformación pseudocircular (SC) y 10 con deformación planolámbdica (PL)-.

Se relevaron ocho rasgos no métricos siguiendo las recomendaciones de Buikstra y Ubelaker (1994): foramen supraorbitario (Forsup), foramen infraorbitario múltiple (Forinf), foramen parietal (Forpar), sutura infraorbitaria (Sutinf), canal del hipogloso dividido (Canhip), canal condilar (Canco), puente pterigo espinoso (Pupte) y huesos lambdoideos (Huelam). Cada rasgo fue registrado como presente/ausente y en aquellos casos de expresión bilateral se consideró únicamente el lado izquierdo. Con el objetivo de evaluar el error intra e interobservador se realizó un diseño experimental de bloques completos aleatorios con medidas repetidas (Cochran y Cox, 1997). Cuatro observadores sin experiencia previa en el registro de este tipo de información relevaron las variables en tres oportunidades espaciadas a intervalos regulares de siete días. Considerando que uno de los objetivos fue evaluar el efecto del entrenamiento sobre el error de observación, dos de los observadores (observador entrenado 1 y observador entrenado 2) recibieron un entrenamiento consistente en la explicación de las definiciones de las variables y en la observación de un conjunto de cráneos. En cada una de las series el orden de los cráneos se estableció al azar. La aleatorización de los bloques y los intervalos de tiempo entre las series aseguran la independencia de las observaciones.

Las comparaciones intra e interobservador fueron realizadas mediante el cálculo del índice Kappa (k) (Cohen, 1960) y la prueba de McNemar (McNemar, 1947). El primero indica el grado de acuerdo que existe por encima del esperado por azar entre pares de observaciones; en tanto la prueba de McNemar permite analizar la homogeneidad marginal (i.e. las proporciones de desacuerdo) y detectar errores sistemáticos en el registro de las variables (Bernal et al., 2004). Se calculó la potencia estadística para la prueba de McNemar de manera de evaluar el alcance de los resultados obtenidos. Para el índice Kappa no se estimó la misma dado que no se emplea con fines inferenciales, sino como una medida del grado de acuerdo entre pares de observaciones (Donner, 1998; Tooth y Ottenbacher, 2004). Asimismo, con el fin de poner a prueba las hipótesis formuladas más arriba se calculó la distancia biológica entre la muestra con deformación pseudocircular y la que presenta deformación planolámbdica mediante la Medida Media de Divergencia (MMD) propuesta por C.A.B. Smith

(Grewal, 1962). La MMD fue calculada sobre los resultados de la última serie de cada observador utilizando la transformación de Freeman-Tukey (Green y Suchey, 1976). La significación de las distancias se determinó con el cálculo del desvío estándar (Sjøvold, 1973). Posteriormente, se realizó un análisis de escalamiento multidimensional (Kruskal, 1964a; b) sobre las distancias MMD. Por último, se comparó mediante una prueba de Mantel (Mantel, 1967; Smouse y Long, 1992; Manly, 1994; Waddle, 1994) la matriz de distancias MMD con dos matrices que describen las expectativas generadas por las dos hipótesis formuladas anteriormente (Tabla 1).

Tabla 1

Matrices de 0 y 1 que describen las hipótesis A (a) y B (b) formuladas para comparar con las matrices de las medidas de distancia MMD

a

	SCe1	PLe1	SCne1	PLne1	SCe2	PLe2	SCne2	PLne2
SCe1	0							
PLe1	1	0						
SCne1	0	1	0					
PLne1	1	0	1	0				
SCe2	0	1	0	1	0			
PLe2	1	0	1	0	1	0		
SCne2	0	1	0	1	0	1	0	
PLne2	1	0	1	0	1	0	1	0

b

	SCe1	PLe1	SCne1	PLne1	SCe2	PLe2	SCne2	PLne2
SCe1	0							
PLe1	1	0						
SCne1	1	1	0					
PLne1	1	1	1	0				
SCe2	0	1	1	1	0			
PLe2	1	0	1	1	1	0		
SCne2	1	1	1	1	1	1	0	
PLne2	1	1	1	1	1	1	1	0

RESULTADOS Y DISCUSION

En la Figura 1 se presentan para cada uno de los observadores los valores del índice Kappa entre la primera y la segunda y entre la segunda y la tercera observación o serie. Los resultados obtenidos indican que hay un incremento en la consis-

tencia intraobservador a lo largo de la experiencia (Figura 1) tanto en los observadores entrenados (e1 y e2) como en los no entrenados (ne1 y ne2). Solo una variable, el puente pterigo espinoso, no presentó un aumento de la concordancia a través de las observaciones (Figura 1). Según los resultados de la prueba de McNemar sólo en dos casos hubo errores sistemáticos, ambos corresponden a comparaciones entre la primera y segunda observación (Figura 2). En los dos casos en que las diferencias fueron significativas la potencia de la prueba fue superior a 0,80. Debe señalarse que los valores ausentes en la Figura 2 corresponden a aquellos pares de observaciones en los que al menos una de ellas era constante y por lo tanto no fue posible realizar la prueba de McNemar.

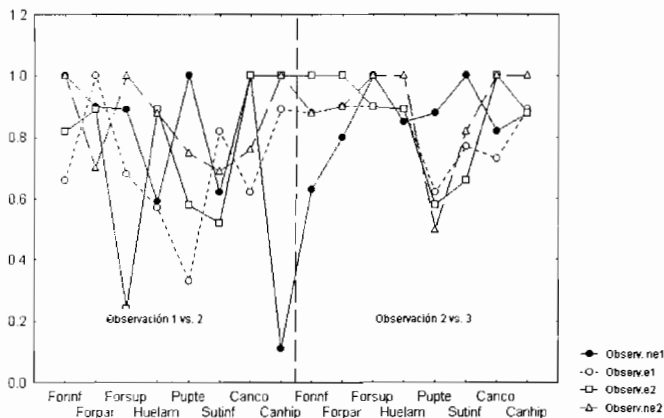


Figura 1

Resultados del índice Kappa calculado sobre las series 1 vs. 2 y 2 vs. 3 de los cuatro observadores.

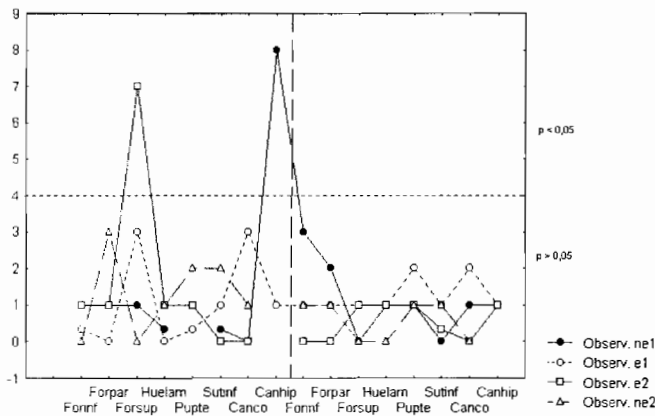


Figura 2

Resultados de la prueba de McNemar efectuada sobre las series 1 vs. 2 y 2 vs. 3 de los cuatro observadores.

En la Figura 3 se observan los resultados del índice de concordancia Kappa entre observadores (ne1-e1; ne1-e2; e1-e2; ne1-ne2; e2-ne2; e1-ne2) para la primera, segunda y tercera observación y para las ocho variables relevadas. Los valores del índice Kappa de la última serie oscilan entre 0,4 y 0,7 en la mayor parte de las comparaciones interobservador (Figura 3). Sin embargo, los resultados difieren según los rasgos considerados, los huesos lambdaideos y el puente pterigo espinoso presentan valores bajos del índice Kappa en las tres observaciones, pero los valores correspondientes al foramen superior se incrementan en las dos últimas observaciones (Figura 3). La concordancia interobservador es más baja que la intraobservador y la inconsistencia se incrementa a lo largo de las series, aún entre los dos observadores entrenados. En la Figura 4 se presentan los resultados de la prueba de McNemar. Se encontraron diferencias significativas en 17 de las comparaciones realizadas. La potencia estadística fue superior a 0,8 para los valores de McNemar mayores a 5 (8 comparaciones). Entre los observadores no entrenados no se detectaron sesgos sistemáticos en el registro de las variables en ninguna de las series, en tanto que los observadores entrenados presentaron sesgos sistemáticos en una variable en las dos primeras observaciones y en tres de las variables en la última serie (Figura 4).

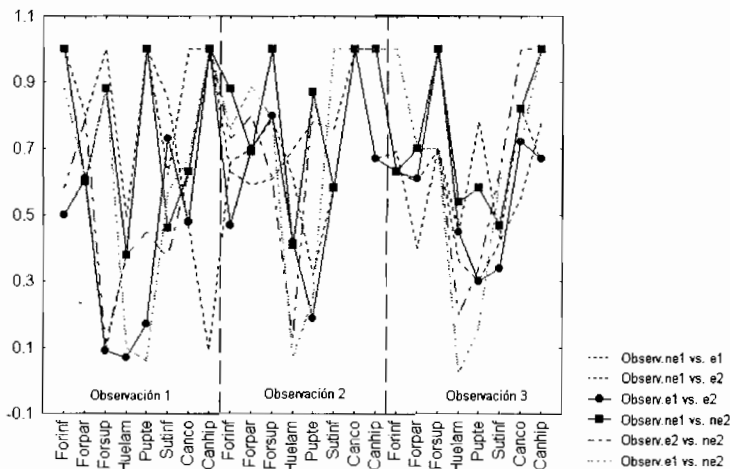


Figura 3

Resultados del índice de Kappa calculado entre los cuatro observadores para las tres series registradas.

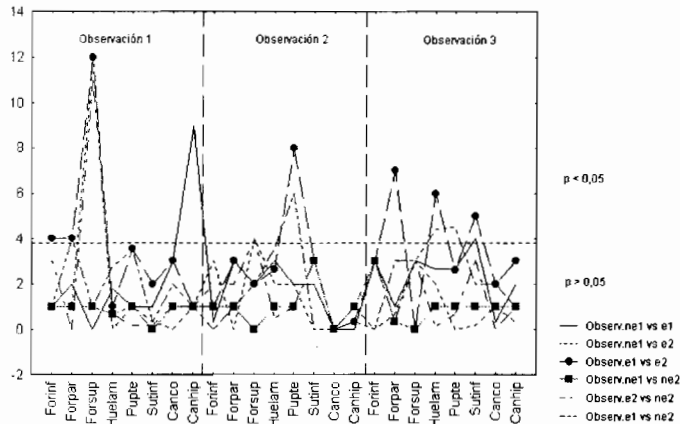


Figura 4

Resultados de la prueba de McNemar efectuada entre los cuatro observadores para las tres series registradas.

El elevado grado de error interobservador encontrado en el registro de los rasgos no métricos puede deberse a la falta de una estandarización adecuada de los procedimientos de registro. Como señalan Molto (1979) y Saunders (1989) no existen estándares que ilustren el grado de expresión, comparables a los disponibles para los rasgos no métricos dentales, por lo tanto, las manifestaciones parciales de los rasgos son incomparables entre observadores y están sujetas a un error considerable. Esta afirmación es válida aún cuando los rasgos son registrados en términos de presencia/ausencia ya que el observador debe decidir a partir de qué nivel de expresión considerará como presente cada rasgo.

No se encontraron diferencias significativas en las distancias MMD entre las muestras con deformación pseudocircular y planolámbdica observadas por un mismo observador, ni entre las muestras relevadas por distintos observadores. Esto puede explicarse por el número de especímenes en cada submuestra. Sin embargo, en el análisis de escalamiento multidimensional (Figura 5) se observa que en algunos casos (v.g. SCne2 y PLel) las diferencias observadas, si bien no son significativas, corresponden a diferencias entre los observadores. Estos resultados son confirmados por la prueba de Mantel (10000 remuestreos) ya que la correlación entre la matriz de distancias MMD y las matrices que representan las hipótesis A y B es baja y no significativa (Hipótesis A: $r = -0,16451$, $p = 0,1650$; Hipótesis B: $r = 0,02356$, $p = 0,5505$). Es decir, las diferencias encontradas entre las muestras medidas por los cuatro observadores no son atribuibles a diferencias morfológicas, ni al entrenamiento de los observadores.

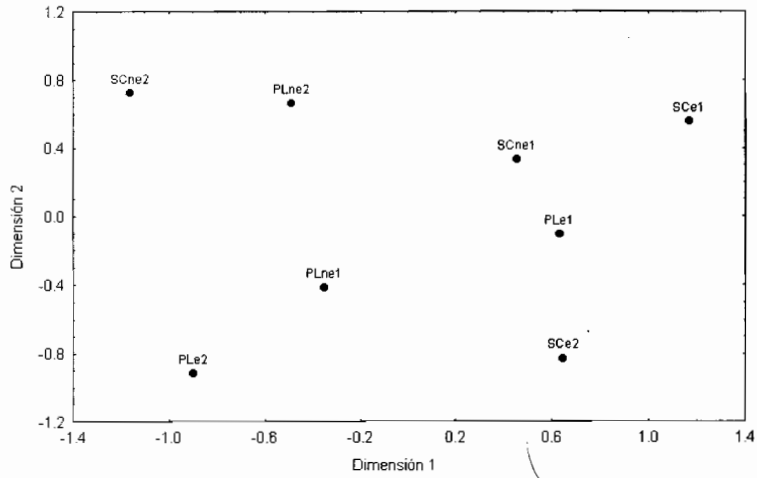


Figura 5

Resultados del análisis de escalamiento multidimensional efectuado sobre las distancias MMD calculadas sobre la última serie de los cuatro observadores (PLe= muestra de individuos con deformación planolámbdica medida por un observador entrenado; PLne= muestra de individuos con deformación planolámbdica medida por un observador no entrenado; SCe= muestra de individuos con deformación pseudocircular medida por un observador entrenado; SCne= muestra de individuos con deformación pseudocircular medida por un observador no entrenado).

CONCLUSIONES

Los resultados obtenidos indican que la aplicación de diseños experimentales permite incrementar la consistencia intraobservador en el registro de los rasgos no métricos. Por el contrario, el error interobservador se incrementa a través de las sucesivas series entre todos los observadores. Estos resultados sugieren que no es conveniente emplear datos relevados por otros investigadores en los análisis de relaciones biológicas. La baja consistencia entre los observadores y la presencia de sesgos sistemáticos modifican las frecuencias de los rasgos en las muestras y por lo tanto alteran las medidas de distancia calculadas a partir de ellas. Debe considerarse, además, que cuando se emplean muestras de tamaño reducido, como las utilizadas generalmente en los estudios bioarqueológicos, pequeñas inconsistencias entre los observadores pueden alterar considerablemente las frecuencias observadas.

AGRADECIMIENTOS

Al Dr. Héctor M. Pucciarelli por facilitar el acceso a las colecciones del Departamento Científico de Antropología del Museo de La Plata (FCNyM. UNLP) a su cargo. Esta investigación fue realizada con fondos del Subsidio de Inicio de Carrera N° 14116-111 de la Fundación Antorchas, dirigido por el Dr. G. Barrientos.

BIBLIOGRAFIA CITADA

Barrientos G y Perez SI (2004) Was there a population replacement during the late middle Holocene in the southeastern Pampas of Argentina? Discussing its archaeological evidence and its paleoecological basis. En Zárata M, Neme G y Gil A (eds): Quaternary International (Special Volume). Ocupaciones Humanas Durante el Holoceno Medio en el Extremo Sudamericano. INQUA, Pergamon Press. En prensa.

Bernal V, González P, Perez I y Del Papa M (2004) Evaluación del error intraobservador en bioarqueología. *Intersecciones en Antropología* 5:129-140.

Berón M y Baffi EI (2003) Procesos de cambio cultural en los cazadores-recolectores de la provincia de La Pampa, Argentina. *Intersecciones en Antropología* 4:29-43.

Berry CA y Berry RJ (1967) Epigenetic variation in human cranium. *J. Anat.* 101:361-379.

Buikstra J y Ubelaker D (1994) Standards for data collection from human skeletal remains. *Arkansas Archaeological Survey Research Series* 44.

Cochran WG y Cox GM (1997) *Diseños Experimentales*. México, Editorial Trillas.

Cohen J (1960) A coefficient of agreement for nominal scales. *Educational Psychology Measurement* 20:37-46.

Corrucini RS (1974) An examination of the meaning of cranial discrete traits for human biological studies. *Am. J. Phys. Anthropol.* 40:425-446.

Donner A (1998) Sample size requirements for the comparison of two or more coefficients of inter-observer agreement. *Stat. Med.* 17:1157-1168.

Finnegan M y Rubison RM (1980) The assessment and analysis of interobserver error in non metric cranial studies. *Am. J. Phys. Anthropol. Suppl.* 52:226.

González-José R, Dahinten S y Hernández M (2001) The settlement of Patagonia: A matrix correlation study. *Hum. Biol.* 73:233-248.

Green RF y Suchey JM (1976) The use of inverse sine transformation in the analysis of non-metric cranial data. *Am. J. Phys. Anthropol.* 45:61-68.

Grewal MS (1962) The rate of divergence in the C57BL strain of mice. *Genetics Research* 3:226-237.

Gualdi-Russo E, Tasca MA y Brasili P (1999) Scoring of nonmetric traits: a methodological approach. *J. Anat.* 195:543-550.

Hanihara T e Ishida H (2001a) Frequency variations of discrete cranial traits in major human populations. I. Supernumerary ossicle variations. *J. Anat.* 198:689-706.

Hanihara T e Ishida H (2001b) Frequency variations of discrete cranial traits in major human populations. II. Hypostotic variations. *J. Anat.* 198:707-725.

Hanihara T, Ishida H y Dodo Y (2003) Characterization of biological diversity through analysis of discrete cranial traits. *Am. J. Phys. Anthropol.* 121:241-251.

Hooton E (1918) On certain eskimoid characters in Icelandic skulls. *Am. J. Phys. Anthropol.* 1:58-62.

Konigsberg LW, Kohn LAP y Cheverud JM (1993) Cranial deformation and nonmetric trait variation. *Am. J. Phys. Anthropol.* 90:35-48.

Kruskal JB (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrics* 29:1-27.

Kruskal JB (1964b) Nonmetric multidimensional scaling: a numerical method. *Psychometrics* 29:115-129.

Manly BFJ (1994) *Multivariate Statistical Methods*. Ed. Chapman & Hall.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27:209-220.

Molto JE (1979) The assessment and meaning of intraobserver error in population studies based on discontinuous cranial traits. *Am. J. Phys. Anthropol.* 51:333-344.

McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12:153-157.

Ossenberg NS (1976) Within and between race distance in population studies based on discrete traits of the human skull. *Am. J. Phys. Anthropol.* 45:701-716.

Outes FF (1911) Variaciones y anomalías anatomo-antropológicas en los huesos del cráneo de los primitivos habitantes del sur de Entre Ríos. *Rev. Museo La Plata* 18:53-144.

Pardoe C (1991) Isolation and evolution in Tasmania. *Current Anthropology* 32:1-21.

Saunders SR (1989) Nonmetric skeletal variation. En Iscan MY y Kennedy KAR (eds): *Reconstruction of Life from the Skeleton*. New York, Alan R Liss Inc., pp. 95-108.

Sciulli PW (1990) Cranial metric and discrete trait variation and biological differentiation in terminal Late Archaic Ohio: The Duff Site cemetery. *Am. J. Phys. Anthropol.* 82:19-30.

Sjøvold T (1973) The occurrence of minor non-metrical variants in the skeleton and their quantitative treatment for population comparisons. *Homo* 24:204-233.

Smouse PE y Long JC (1992) Matrix correlation analysis in anthropology and genetics. *Am. J. Phys. Anthropol.* 35:187-213.

Tooth LR y Ottenbacher KJ (2004) The statistic in rehabilitation research: an examination. *Arch. Phys. Med. Rehabil.* 85:1371-1376.

Waddle DM (1994) Matrix correlation test support a single origin for modern humans. *Nature* 368:452-454.

Zegura SL (1975) Taxonomic congruence in Eskimoid populations. *Am. J. Phys. Anthropol.* 43:271-284.