ARTÍCULO DE INVESTIGACIÓN

# La estimación del error en métodos cuantitativos para identificación humana: un experimento con las arcadas dentales

**Mirsha Quinto-Sánchez[1]✉, Sadahi García[2], Andrea Nares[3], Carolina Hernández[2], Cynthia San Juan[2], Stefano Huitrón[2], Chantal Loyzance[1], Ivet Gil-Chavarría[1]**

[1]Ciencia Forense, Facultad de Medicina, Universidad Nacional Autónoma de México.
[2]Facultad de Odontología, Universidad Nacional Autónoma de México.
[3]Facultad de Medicina, Universidad Nacional Autónoma de México.

✉ mirsha@cienciaforense.facmed.unam.mx

Datos del artículo:

Cita: Quinto-Sanchez M, García S, Nares A, Hernández C, San Juan C, Huitron S, Loyzance C, Gil-Chavarría I. 2022. La estimación del error en métodos cuantitativos para identificación humana: un experimento con las arcadas dentales. Revista Digital de Ciencia Forense. 1(1): 1-22 pp.

Editor: Vicente Torres.

Recibido: 21 septembre 2021.

Aceptado: 21 febrero 2022.

Publicado: 28 abril 2022.

Financiamiento: Conacyt PN-2016-2518.

## Resumen

Los efectos del error de medición (EM) intra e inter observador en los datos empleados en métodos cuantitativos en la identificación humana, deben ser evaluados para evitar la aparición de errores estadísticos en la interpretación de datos, y la consecuente generación de falsos positivos o negativos. El objetivo fue evaluar el error intra e inter en nueve observadores en un protocolo de 20 landmarks, caracterizando su fluctuación en cuatro iteraciones independientes. Se contrastó la hipótesis de que el patrón del error de medición disminuye en función del tiempo y en relación con la formación y/o área académica del observador (experiencia). Para ello, se fenotiparon 139 fotografías de alta resolución de modelos dentales de 45 individuos de la Colección Odontológica Nacional de Ciencia Forense de la UNAM. Después, mediante el uso la morfometría geométrica y estadística multivariada se analizó la variación general de la muestra con un diseño experimental anidado por observadores e iteraciones. Los resultados fueron contrarios a lo esperado, los datos no muestran una disminución del error en función del tiempo. También, que el error inter fue mayor que el intraobservador, como otros estudios han reportado. La mayor frecuencia de error fue entre observadores por la secuencia de posicionamiento de landmarks. Con base en los resultados se recomiendan generalidades para evitar la aparición de error en estudios morfogeométricos, la principal es la inclusión de reportes de error intra e inter observador en los peritajes que usen mediciones o digitalizaciones.

**Palabras clave:** Error de medición, morfometría geométrica, identificación humana.

## Abstract

The effects of intra and inter-observer measurement error (EM) in the data used in quantitative methods in human identification must be evaluated to avoid the appearance of statistical errors in the interpretation, and the consequent generation of false positives or negatives. The objective was to evaluate the intra and inter-error in nine observers in a protocol of 20 landmarks, characterizing its fluctuation in four independent iterations. The hypothesis that the measurement error pattern decreases as a function of time and in relation to the observer's training and / or academic area (experience) was tested. For this, 139 high resolution photos of dental casts from 45 individuals from the National Forensic Science Dental Collection of UNAM were phenotyped. Then, by using geometric morphometrics approach and multivariate statistics, the general variation of the sample was analyzed with a nested design by observers and iterations. The results show that contrary to expectations, the data does not show a decrease in error as a function of time. Also, that the interobserver error was greater than the intraobserver one, as other studies have reported. The highest frequency of error was between observers due to the sequence of positioning of landmarks. Based on the results, generalities are recommended to avoid the appearance of error in geometric morphometric forensic studies, the main one being the inclusion of intra- and inter-observer error reports in expert opinions that use measurements or digitizations.

**Keywords:** measurement error, geometric morphometric, human identification.

## Introducción

La observación fundamental de la biología es la morfología (1). En años recientes el estudio morfológico ha sido encabezado por la morfometría, que se encarga de la cuantificación de las formas biológicas. Dicho proceso conlleva una técnica de medición, misma que impone una limitación práctica a lo que puede estudiarse: la capacidad del observador de realizar mediciones precisas, es decir, su error de registro (2).

El error de medición (EM) ha sido definido como el porcentaje de la variación fenotípica atribuible al error de medida (3). Sin embargo, esta definición atribuye la variación a un factor extrínseco o no inherente a ella, por lo que se considera incorrecta (2). Una definición más aceptada es que el EM "es la desviación del resultado en una medición al valor verdadero de ésta" (3). En una caracterización del EM se busca que: la precisión, que es la dispersión del conjunto de valores obtenidos de mediciones repetidas de una magnitud, y la exactitud, que es cuán cerca del valor real se encuentra el valor medido (Figura 1a) se distribuyan cercanos o tendientes a cero.

El EM puede ser caracterizado en varios niveles: dentro de observadores (repetibilidad), entre observadores (reproducibilidad) y entre diferentes métodos (reproducibilidad). Para profundizar en lo anterior, la repetibilidad es la cercanía entre los resultados de mediciones sucesivas de la misma magnitud por medir, efectuadas en las mismas condiciones de medición (Figura 1b). Las condiciones de repetibilidad incluyen: un mismo procedimiento de medición, el mismo observador, el mismo instrumento de medición utilizado en las condiciones y un mismo lugar y repetición dentro de un periodo de tiempo corto (4,5). La reproducibilidad es la cercanía entre los resultados de las mediciones de la misma magnitud por medir, efectuada bajo mediciones diferentes (Figura 1b) como pueden ser: el principio de medición, el método de medición, el observador, el instrumento de medición, el patrón de referencia, el lugar y las condiciones de uso y el tiempo (4). Es así como la repetibilidad y la reproducibilidad permiten la plasticidad de cambio de condiciones para generar programas de investigación prolíficos respecto de la caracterización del error que en antropología biológica y forense son bastos (6–14).

El EM, entonces, es una constante en los estudios de caracteres morfométricos (15–17) que se puede circunscribir bajo los siguientes supuestos generales: 1) el resultado de una medición es sólo una aproximación o estimación del valor de la cantidad específica sometida a medición, es decir, magnitudes de medición, y por lo tanto el resultado es completo sólo cuando se acompaña de una evaluación cuantitativa de su incertidumbre (18), 2) la sofisticación de los instrumentos de medición no implica la eliminación de error en las mediciones, por el contrario el error siempre se encontrará presente y reducirlo es un elemento importante del quehacer morfométrico (17), 3) la caracterización del error es una constante en todo estudio cuantitativo que implique la medición

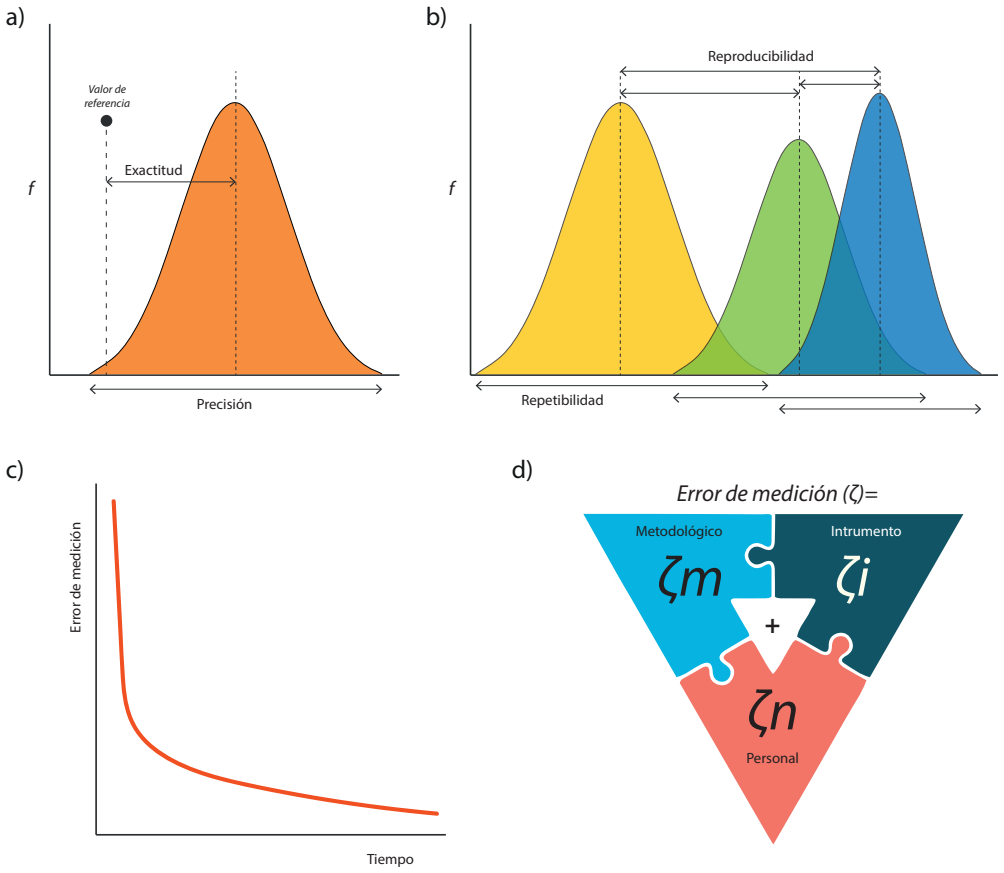


**Figura 1**. Fundamento teórico del estudio de error en estudios cuantitativos. A) relación estadística entre el concepto de precisión y exactitud. B) relación entre la repetibilidad y la reproducibilidad en términos estadísticos con tres grupos de medidas. C) dogma del error de medición donde éste disminuye en función del tiempo (experiencia) del observador y su entrenamiento respectivo. D) Modelo general del error en la morfometría geométrica según Arnqvist y Martensson (24), que es la sumatoria del ζm es el error metodológico, ζi el error del instrumento y ζn el error personal.

de alguna estructura , y 4) se ha constatado que los protocolos utilizados por la morfometría geométrica presentan menos problemas con el error respecto de los empleados por la morfometría tradicional (17).

En las ciencias forenses, específicamente en antropología forense, existe la idea generalizada de que el error disminuye en función del tiempo (Figura 1c), como un efecto directo de la experiencia del observador, aquí este fenómeno lo denominamos "dogma del EM". Comúnmente, durante la presentación de peritajes o incluso en el desarrollo de métodos cuantitativos (que se pueden extender a los morfoscópicos o visuales) se omite el reporte del EM, porque se asume que, por experiencia, su error tendería a ser bajo o nulo. Incluso, se asume que el EM entre pares en algún equipo forense es bajo y constante a través del tiempo. Esta práctica esta alejada de toda praxis basada en evidencia (19–22) y es por dicho planteamiento que nace el presente trabajo, la caracterización

del EM es un requerimiento estándar para cualquier laboratorio antropológico forense, extensible a otras áreas de las ciencias forenses.

El informe para el fortalecimiento de las ciencias forenses en EEUU (23), estableció la necesidad de programas científicos para examinar el error o sesgo de las observaciones realizadas por peritos en diversos procedimientos forenses, proponiendo la conformación de validación de mediciones, repetibilidad-reproducibilidad de protocolos generales en ciencia forense. Como un punto específico se recomienda la recopilación y análisis de datos forenses: exactitud y tasas de error del análisis forense, fuentes de sesgo potencial y error humano en la interpretación por expertos forenses y pruebas de aptitud de expertos forenses, dando preponderancia a los modelos basados en evidencia y haciendo generalización de la práctica de estimación del EM en las ciencias forenses.

Para realizar dicha caracterización existe un modelo teórico general del EM que Arnqvist y Martensson (24), sintetizan, siguiendo a Rabinovich (25), que establece que:

$$EM(\zeta)=\zeta m +\zeta i +\zeta n \quad [1]$$

Donde, ζm es el error metodológico, ζi el error del instrumento y ζn el error personal. Como se observa en el modelo, cada componente es aditivo y se subdivide como sigue (Figura 1d): el error metodológico, en preparación del espécimen, reducción dimensional (coplanaridad) y preparado de datos; el error instrumental, en distorsión óptica cuando se captura el individuo, distorsión óptica al reproducir el individuo, distorsión digital en la captura y reproducción del individuo; finalmente el error personal, que es el error cometido por el o los observadores en un banco de datos (17).

Taylor y Kuyatt (18), recomiendan el uso de estadísticas como ANOVA o correlaciones en la evaluación de estos efectos o bien vía coeficiente de correlación intraclase (6). Estadísticamente, siguiendo las definiciones de Rohlf (26), un estimador fenotípico será insesgado cuando para todos los valores de la muestra, el valor esperado de la estimación (el valor medio de la estimación) es igual al valor verdadero del parámetro, esto es:

$$E(\grave{q})=q \quad [2]$$

Donde $\grave{q}$ es una estimación del parámetro q (26). De lo contrario, se dice que el estimador está sesgado. El desvío o error, entonces, es la diferencia de:

$$E(\grave{q})-q \quad [3]$$

La magnitud promedio de error a menudo se mide como el error cuadrático medio, MSE (por su significado en inglés, mean square error), el promedio de la desviación al cuadrado de una estimación de su valor real:

$$MSE\grave{}q = E(\grave{}q - q)^2 = s^2 + (E(\grave{}q) - q)^2 \quad [4]$$

Este estadístico proporciona una medida de la exactitud en lugar de sólo la precisión, ya que es una función de la proximidad de las estimaciones a la media real en lugar de la media de la muestra (26).

Las formas para minimizar la contribución del EM en la muestra son diversas, sin embargo, algunas pautas generales ayudan a resolver el problema. Una ruta comúnmente empleada es que los pasos, procesos, tratamientos, estandarizaciones y correcciones desde el individuo original a los datos morfogeométricos deberán ser los menos posibles (17). Es importante señalar que algunas comparaciones de métodos de medición morfométrica manual y digital, muestran que estos últimos presentan mayor precisión y están menormente influenciados por factores que aumentan el EM (27,28), y más aún cuando éstos son realizados en tomografías (8), por lo que la morfometría geométrica, implementando el método de ANOVA (e.g. de una vía o Procrustes ANOVA), deberá presentar un menor umbral al EM, menor al 2% que se establece como umbral clásico y aceptado en morfometría (8,29). Luego entonces, la práctica morfogeométrica deberá estar orientada a amortiguar el error de medición (EM) mediante protocolos robustos de medición, pericia del investigador y correcta implementación de modelos estadísticos.

El objetivo de presente trabajo es evaluar el EM dentro y entre diversos observadores (n=9) en un set de 20 puntos craneofaciales (landmarks), caracterizando su fluctuación en cuatro iteraciones independientes, estableciendo si el patrón del error de medición disminuye en función del tiempo en relación con la formación y/o área académica del observador y su *expertise*. Con ello se establecen recomendaciones generales para la determinación del error en ciencias forenses. Se espera, como hipótesis generalizada, que las mediciones realizadas sean repetibles entre observadores, es decir que los estimadores de repetibilidad y error inter e intra observador sean insesgados ( $E(\grave{}q) = q$), dentro de los estándares establecidos para la morfometría (8,29); es decir, que error intra e inter observador se reduzca en función del tiempo, siendo más preciso y logrando una meseta de error en la última iteración realizada.

## Material y Método

### *La muestra*

Se emplearon 139 fotografías de alta resolución de modelos dentales pertenecientes a 45 individuos de la Colección Odontológica Nacional de Ciencia Forense de la Facultad de Medicina de la Universidad Nacional Autónoma de México. Para cada individuo, se tenían, por lo general, los moldes dentales para la maxila y mandíbulas, pre y post-tratamiento lo que suma cuatro moldes por

persona. Sin embargo, al ser un archivo sin uso clínico, algunos individuos no presentaban todos los moldes, motivo por el cual el total no es 180 fotografías. Específicamente, dos individuos presentaron un molde, 10 individuos dos moldes, tres individuos tres moldes; así que existe discrepancia entre el número de individuos y las fotografías realizadas. Los modelos fueron donados de forma anónima por la División de Estudios de Posgrado e Investigación de la Facultad de Odontología, UNAM, por lo que no se requirió consentimiento informado.

### *Fenotipado*

Para capturar la forma del arco dental se aplicó un protocolo de fotografía científica sobre cada modelo de yeso. Las fotografías se obtuvieron con una cámara Nikon D4 y un lente fijo AF Nikkor 50mm f / 1.8D con un filtro UV de 52mm (Figura 2a). Esta lente fue elegida en base a estudios previos por presentar el factor de deformación esférico o de borde más bajo (30). Se empleó una mesa de reproducción fotográfica para el montaje, el centroide de la cámara se fijó ortogonalmente a 0.5 m de altura, que se mantuvo constante entre tomas. Para su referencia métrica, se colocaron dos escalas cerca del modelo, una metálica contigua al área de enfoque y otra de plástico en la parte inferior de la imagen. La escala de las fotografías se realizó utilizando la escala de metal que estaba en el área de enfoque de la cámara. La cámara fue parametrizada con los siguientes valores: ISO -100, f / 5.6, y exposición compensada para subexponer un paso o nivel. Estos valores dan un campo de profundidad con un límite cercano de 48.5 cm y un límite lejano de 51.6 cm, esto es un 3.06 cm de nitidez que es superior para las longitudes de los elementos dentales en estudio. Finalmente, si bien se sacaron fotografías en formato *.raw, para el presente estudio se emplearon las almacenadas en formato *.jpg, con una resolución de 300 píxeles por pulgada en horizontal y 300 píxeles por pulgada en vertical (Figura 2b).

Después, se colocaron un conjunto de 20 *landmarks* (Figura 2c) sobre las fotografías basado en el protocolo de Kieser (31). La delimitación fue sobre las cara incisales (en anteriores) y oclusales (en posteriores), iniciando por el lado derecho de la fotografía con la cara distal del molar más posterior (*landmarks* 1,2,3), y siguiendo con los puntos de contacto dados por la superposición de los principales grupos de dientes; es decir, molares-premolares (*landmarks* 4,5), premolares-canino (*landmarks* 6,7), canino-incisivos (*landmarks* 8,9) y punto de contacto o línea media de incisivos centrales (*landmarks* 10, 11), siguiendo la secuencia ahora de anterior a posterior del lado izquierdo (ver Figura 2c). En el caso de ausencia dentaria, se colocaron puntos en el centroide de la cara oclusal o incisal o el punto de contacto del grupo de dientes previamente descrito. El fenotipado se realizó utilizando el software TPSDIG (32). Cada sesión de digitalización se llevó acabo con un mes de intervalo, requiriendo a los observadores realizarlo durante la mañana del primer día de cada semana. Además, el protocolo se discutió de forma general entre los observadores y se apoyó siempre con una imagen de referencia anatómica de los *landmarks*

(Figura c).

Este proceso fue realizado por los nueve observadores con formación en distintas disciplinas y con diferentes grados de especialidad, entre estos: medicina [1], antropología [3], odontología [4] e ingeniería [1]. Se realizaron cuatro sesiones independientes de fenotipado con un mes de diferencia entre ellas, con el fin de poder ver la tendencia de evaluación a través del tiempo. Las coordenadas *landmarks* resultantes del presente estudio están disponibles (laofunam.com/data) para metaanálisis.

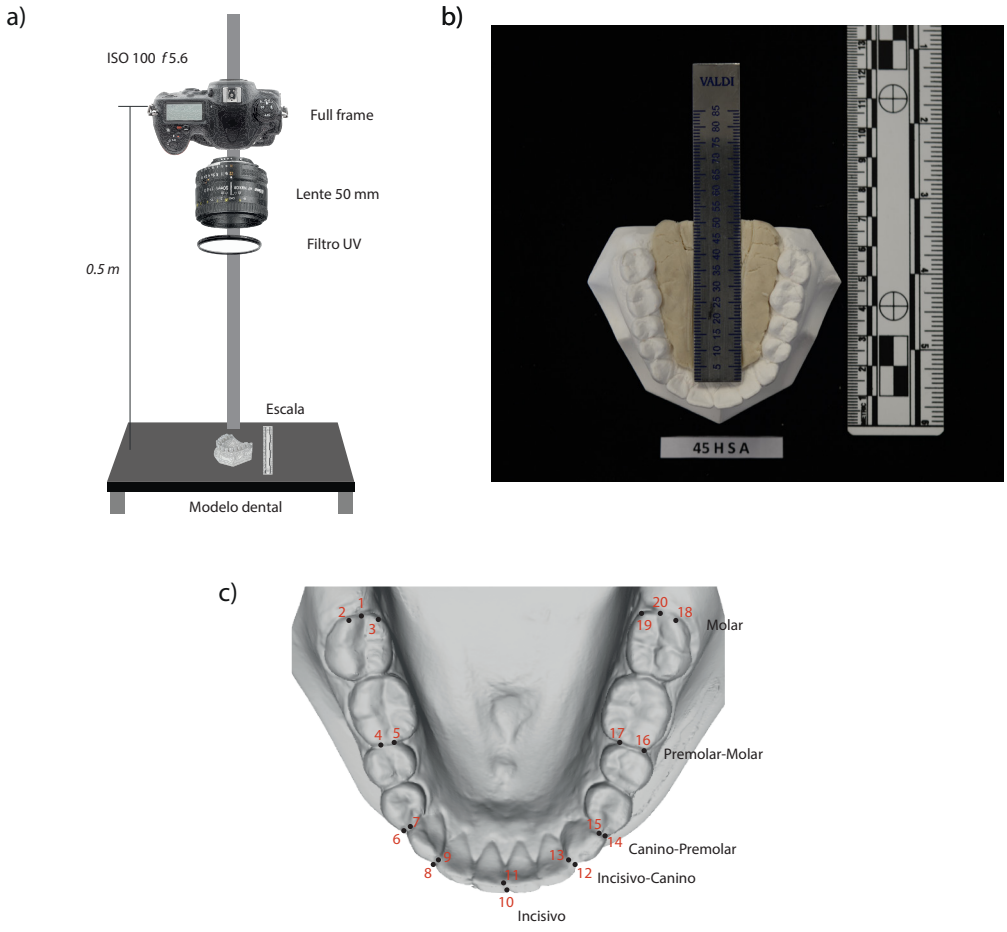


**Figura 2**. Protocolo de fenotipado de la arcada dental anterior por medio de modelos dentales. a) montaje de fotografía científica para la captura de las fotografías ortogonales de los modelos. B) ejemplo del resultado de las fotografías realizadas. C) protocolo de 20 *landmarks* empleados en el estudio en relación con el grupo dental al que pertenecen.

*Diseño experimental*

Sobre las coordenadas de *landmarks* se integró a través de la plataforma MorphoJ (33), un archivo único con todas las digitalizaciones realizadas. Sobre este conjunto de datos se realizó un análisis generalizado Procrustes (GPA, 34–36) con el fin de generar un morfoespacio invariante a los efectos de traslación, rotación y escala (34). Todo el procedimiento se realizó de forma generalizada y por observador en un sistema de anidación de datos. Después, se buscó aquellos individuos atípicos en la muestra (*outliers*) calculando la distancia de cada individuo a la forma promedio (*average shape*), lo que proporciona un indicador de cuán inusual es un individuo en relación con la muestra. Luego, por medio del cálculo de un intervalo $\alpha$=0.01 se estableció un punto de corte para toda la muestra. Así es que se podrá reportar la cantidad de individuos fuera de este intervalo para cada observador, por disciplina y especialización (evaluando si existen diferencias significativas entre éstas últimas).

Posteriormente, se generó una matriz de varianza covarianza, como paso previo para realizar un análisis de componentes principales (ACP) como una forma de caracterizar la covariación en los datos. Sobre este morfoespacio resultante se registró la varianza total por observador e iteración, dicho valor fue multiplicado por 100 para mejorar la escala. Además, se calculó una media para todas las iteraciones por observador, un intervalo $\alpha$=0.01 y fue indicada la varianza total de la muestra, todo ello resumido mediante un gráfico de puntos, donde se evaluó visualmente si la tendencia de las iteraciones era o no disminuir en función del tiempo (iteraciones) y si estas eran diferentes entre disciplinas.

Después, se evaluó el error general cometido por observador siguiendo el protocolo de Singleton (37). Para ello se calculó un ANOVA de una vía para cada landmark por observador e iteración, donde los valores del error cuadrático medio (RMSE) fueron analizados como un indicador del error intra-observador (38). El error intraobsevador no ha de exceder 0.33 mm o 2%, mientras que las medidas más imprecisas no deben exceder el 5% de error (29).

Posteriormente, se calcularon dos análisis de variación canónica (CVA), uno por observador y otro por observador-iteración sobre las coordenadas de Procrustes con 10,000 permutaciones para significar las diferencias entre grupos. También, como resultado del CVA se obtuvieron matrices de distancias de Mahalanobis y Procrustes que fueron exportadas para el análisis posterior. De igual manera, el CVA permitió conocer la morfología de los cambios entre observadores por cada raíz canónica resultante.

Finalmente, las matrices de Mahalanobis fueron sometidas a un análisis de cluster por medio del algoritmo UPGMA. Esto permitió tener un dendograma con 10,000 permutaciones que agrupara a los diferentes observadores e iteraciones en el análisis, esperando que las cuatro iteraciones y todos los observadores coincidieran entre ellos.

# Resultados

## *Determinación de outliers*

Se identificaron 51 individuos fuera del rango de variación normal establecido como el percentil 0.99 (punto de corte 120.36) realizado sobre la distancia de cada individuo a la forma media (*average shape*). Dentro de los observadores se logró identificar que todos ellos presentaron individuos fuera del percentil 99 con 46 casos y únicamente cinco casos con valores extremos (Figura 3). La menor cantidad de individuos fuera del rango fue tres para los observadores 1, 4 y 8 y uno más con 4 individuos (Observador 5). El resto de los observadores tuvieron entre 6 y 11 (Tabla 1). En algún observador se presentaron más de dos valores extremos (Observador 9). De manera general, la mayor cantidad de estos individuos se debieron a errores en la digitalización de los landmarks, por ejemplo, al orden de digitalización y por la apreciación de la colocación de un punto determinado. Además, se constató que había diferencias significativas entre la frecuencia de *outliers* por disciplina y especialización ($p$=9.999e-05).

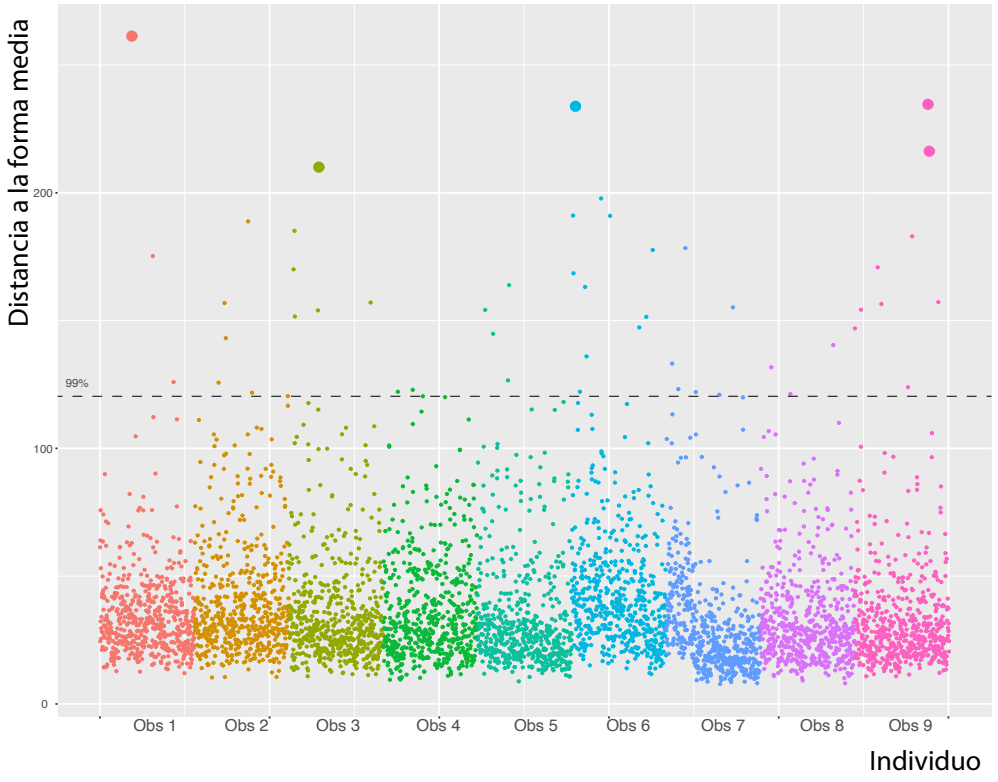


**Figura 3**. Gráfico de *Manhattan* de la distribución de los individuos del estudio respecto de la distancia a la forma media (*average shape*) por cada observador. Sobre estos valores se ha trazado el punto de corte al 99% lo que define los *outliers*. Se identificaron cinco casos extremos (d >200).

**Tabla 1**. Número de valores atípicos (*outliers*), valores de varianza, media, total y especialidad por iteración realizada por cada observador en el diseño anidado.

| | *Outliers* | Itera1 | Itera2 | Itera3 | Itera4 | Media | Total | Especialidad |
|---|---|---|---|---|---|---|---|---|
| Obs 1 | 3 | 0.682 | 0.676 | 0.701 | 0.670 | 0.682 | 0.681 | Medicina |
| Obs 2 | 6 | 0.766 | 0.742 | 0.761 | 0.745 | 0.753 | 0.750 | Odontología |
| Obs 3 | 6 | 0.741 | 0.727 | 0.727 | 0.753 | 0.737 | 0.734 | Ingeniería |
| Obs 4 | 3 | 0.732 | 0.761 | 0.756 | 0.755 | 0.751 | 0.750 | Odontología |
| Obs 5 | 4 | 0.766 | 0.760 | 0.749 | 0.745 | 0.755 | 0.753 | Odontología |
| Obs 6 | 11 | 0.757 | 0.722 | 0.717 | 0.728 | 0.731 | 0.733 | Antropología |
| Obs 7 | 6 | 0.713 | 0.684 | 0.700 | 0.702 | 0.700 | 0.705 | Antropología |
| Obs 8 | 3 | 0.735 | 0.728 | 0.731 | 0.731 | 0.731 | 0.729 | Antropología |
| Obs 9 | 9 | 0.737 | 0.733 | 0.719 | 0.750 | 0.735 | 0.735 | Odontología |

### *Análisis de la varianza de los morfoespacios*

Los valores de varianza total de los morfoespacios permiten generalizar el comportamiento del experimento realizado. De manera general, se obtuvo una varianza total general de 0.744 con una media entre observadores 0.73 (con valores mínimo=0.67, máximo=0.765 entre iteraciones, Tabla 1). En adición, la varianza total entre las diferentes iteraciones por observador muestra que la tendencia general no es la disminución en función del tiempo (Tabla1, Figura 4) ese patrón únicamente se observó en un observador (Observador 5); además, un observador con un patrón de varianza muy baja (Observador 8). El Anova de una vía por observador resultó en diferencias significativas ($F$=16.15 $p$=1.86e-08), la prueba TukeyHSD mostró 11 comparaciones entre observadores que resultaron significativas (Tabla 2), donde el observador 1 fue quien más diferencias entre observadores acumuló con un total de siete.

### *Cálculo del error intra e interobservador*

De manera generalizada, en la aplicación del método Singleton (Tabla 3) resultó en un porcentaje medio de RMSE intraobservador de 0.0028 (mínimo=0.0003, y máximo=0.0078) resultando en diferencias significativas en el landmark 1 del observador 3 ($p$=0.0388), diluyéndose la significación en la prueba de Tukey *pos hoc* ($p$=0.0668, ver suplementario 1 STabla 1). Los valores de error exceden el rango esperado del 5%.

### *Diferencias entre observadores*

Respecto de los valores de RMSE interobservadores (Tabla 4) se tuvo un valor promedio de 0.0059 (mínimo=0.0029, y máximo=0.0114), resultando en

**Tabla 2**. Resultados de la prueba *pos hoc* de Tukey sobre la varianza de los morfoespacios para las comparaciones entre observadores.

| | Comparación | Diferencia | Inferior | Superior | *p* |
|---|---|---|---|---|---|
| 1 | Obs5-Obs1 | 0.00072719 | 4.35E-04 | 1.02E-03 | 0.0000002 |
| 2 | Obs2-Obs1 | 0.0007128 | 4.21E-04 | 1.00E-03 | 0.0000003 |
| 3 | Obs4-Obs1 | 0.00068736 | 3.95E-04 | 9.79E-04 | 0.0000005 |
| 4 | Obs7-Obs5 | -0.0005495 | -8.41E-04 | -2.58E-04 | 0.0000278 |
| 5 | Obs3-Obs1 | 0.0005458 | 2.54E-04 | 8.38E-04 | 0.000031 |
| 6 | Obs7-Obs2 | -0.0005351 | -8.27E-04 | -2.43E-04 | 0.0000425 |
| 7 | Obs9-Obs1 | 0.0005244 | 2.32E-04 | 8.16E-04 | 0.0000585 |
| 8 | Obs7-Obs4 | -0.0005097 | -8.02E-04 | -2.18E-04 | 0.0000907 |
| 9 | Obs8-Obs1 | 0.00048893 | 1.97E-04 | 7.81E-04 | 0.0001689 |
| 10 | Obs6-Obs1 | 0.00048881 | 1.97E-04 | 7.81E-04 | 0.0001695 |
| 11 | Obs7-Obs3 | -0.0003681 | -6.60E-04 | -7.62E-05 | 0.0061237 |

**Tabla 3**. Resultados de método si para el error intraobservador, se presentan los valores de mínimo (min), máximo (max), media y desviación estándar (de) de los valores de RMSE y porcentajes. Abajo se señala el caso que fue significativo (*).

| | | mín | máx | media | de |
|---|---|---|---|---|---|
| Observador 1 | rmse | 0.0010 | 0.0042 | 0.0026 | 0.0009 |
| | % rmse | 1.9884 | 8.0680 | | |
| Observador 2 | rmse | 0.0007 | 0.0056 | 0.0024 | 0.0011 |
| | % rmse | 1.3775 | 11.8546 | | |
| Observador 3* | rmse | 0.0007 | 0.0054 | 0.0027 | 0.0015 |
| | % rmse | 1.3152 | 9.9006 | | |
| Observador 4 | rmse | 0.0006 | 0.0078 | 0.0038 | 0.0022 |
| | % rmse | 0.8409 | 10.3674 | | |
| Observador 5 | rmse | 0.0010 | 0.0036 | 0.0020 | 0.0008 |
| | % rmse | 2.5286 | 8.9120 | | |
| Observador 6 | rmse | 0.0012 | 0.0048 | 0.0027 | 0.0010 |
| | % rmse | 2.2243 | 8.9968 | | |
| Observador 7 | rmse | 0.0010 | 0.0073 | 0.0031 | 0.0015 |
| | % rmse | 1.5265 | 11.7090 | | |
| Observador 8 | rmse | 0.0003 | 0.0069 | 0.0030 | 0.0016 |
| | % rmse | 0.5816 | 11.6211 | | |
| Observador 9 | rmse | 0.0015 | 0.0074 | 0.0035 | 0.0019 |
| | % rmse | 2.1565 | 10.5844 | | |

* $p$=0.0388 *landmark* 10.

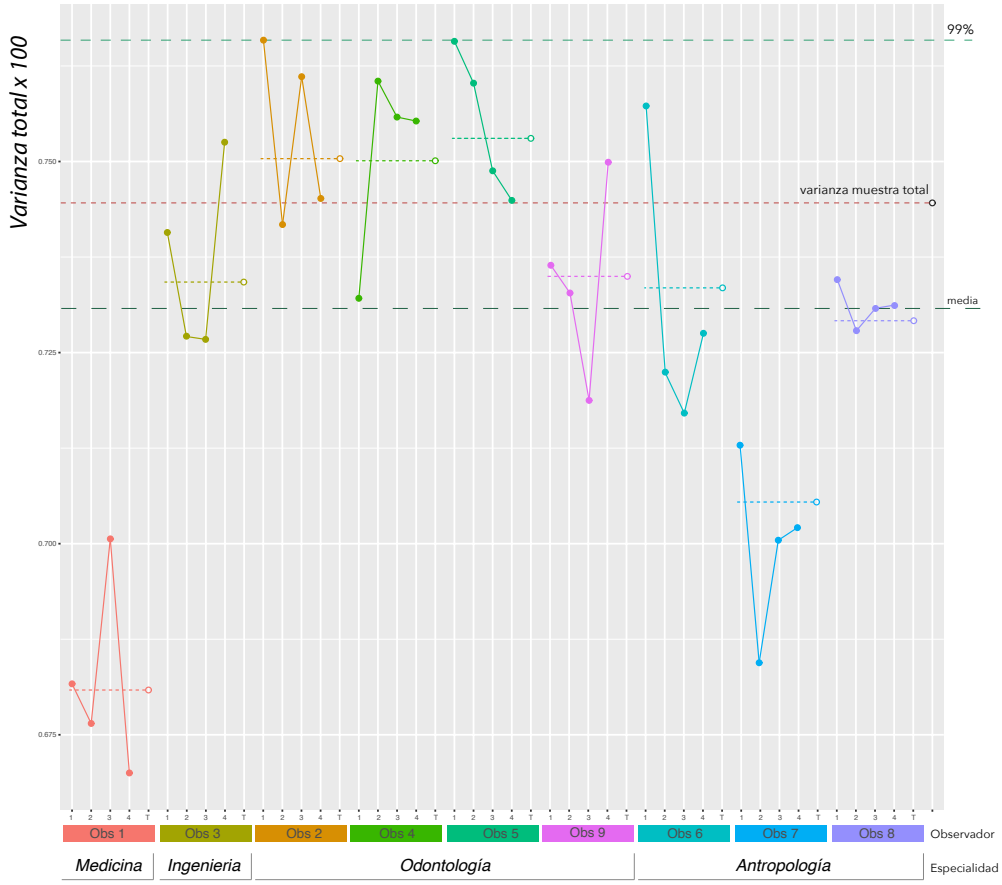


**Figura 4**. Gráfico de puntos de la varianza x100 de cada una de las cuatro iteraciones realizada por observador en el experimento, ordenadas por especialidad (medicina, ingeniería odontología y antropología). En cada observador se presenta el valor de varianza por cada una de las iteraciones, unidas por una línea para denotar la tendencia. Además, se grafica la varianza total por observador (T) con una línea discontinua del color respectivo. Se trazaron tres líneas referencia para: a) la media de todo el experimento (línea discontinua abierta en negro), b) la varianza muestra total (línea discontinua cerrada roja) y c) el intervalo del 99% de las varianzas obtenidas (línea discontinua abierta verde).

diferencias significativas para el *landmark* 10 ($p$=0.0351) y 11 ($p$=0.01729) de la iteración 2, los *landmark* 10 ($p$=7.73e-05) y 11 ($p$=4.26e-06) de la iteración 3; y los *landmark* 1 (p=0.0401), 10 ($p$=0.0313), 11 ($p$=0.0395) y 12 ($p$=0.0467) de la iteración 4 (ver Suplementario 1, STabla 2, 3, 4). Como se puede ver en la tabla 4 todos los valores resultantes de los porcentajes de error están por encima de lo esperado que es un 5 % de error como punto más alto.

Respecto del CVA, se generó un espacio con ocho raíces canónicas para explicar el 100% de la varianza (Tabla 5), los tres primeros ejes explican un total de 77.04% de varianza. Los cambios de forma resultantes permitieron caracterizar que el principal problema de los errores en la digitalización es

**Tabla 4**. Resultados de método Singleton para el error interobservador, se presentan los valores de mínimo (mín), máximo (máx), media y desviación estándar (de) de los valores de RMSE y porcentajes. Abajo se señala el caso que fue significativo (*).

| | | mín | máx | media | de |
|---|---|---|---|---|---|
| Iteracion 1 | rmse | 0.0030 | 0.0091 | 0.0052 | 0.0015 |
| | % rmse | 2.8742 | 8.7010 | | |
| Iteracion 2* | rmse | 0.0032 | 0.0115 | 0.0062 | 0.0021 |
| | % rmse | 2.6299 | 9.3128 | | |
| Iteracion 3* | rmse | 0.0037 | 0.0094 | 0.0062 | 0.0018 |
| | % rmse | 2.9888 | 7.5317 | | |
| Iteracion 4* | rmse | 0.0034 | 0.0113 | 0.0062 | 0.0020 |
| | % rmse | 2.7822 | 9.0868 | | |

* con significación estadística

la secuencia de los puntos, es decir la pérdida de la homología estricta, estos individuos son los *outliers* encontrados en el estudio, reportados más arriba. La matriz de Mahalanobis (Tabla 6) resultó significativa en todas las distancias ($p$=<0.0001) con una media de distancias entre observadores de 2.83, un mínimo de 1.38 y un máximo de 4.75. La mayor distancia entre observadores se presenta entre el observador 2 y el 6, seguidos por el observador 1 y 2 (Tabla 6). Estos datos resultaron consistentes con la métrica de Procrustes donde, si bien algunos valores $p$ disminuyeron ($p$=0.0058 como mínimo) se mantuvo la significación estadística en todas las comparaciones (ver Suplementario 1, STabla 5).

**Tabla 5**. Resultados del análisis de variación canónica (CVA) para los observadores, se presentan los eigenvalores, porcentaje de varianza explicada por componente principal y varianza acumulada.

| | *eigenvalues* | % Varianza | % Acumulada |
|---|---|---|---|
| CV1 | 2.17244244 | 55.902 | 55.902 |
| CV2 | 0.46595603 | 11.99 | 67.893 |
| CV3 | 0.35583403 | 9.156 | 77.049 |
| CV4 | 0.32304543 | 8.313 | 85.362 |
| CV5 | 0.21881024 | 5.631 | 90.992 |
| CV6 | 0.17895961 | 4.605 | 95.597 |
| CV7 | 0.09833095 | 2.53 | 98.128 |
| CV8 | 0.07276055 | 1.872 | 100 |

En el analisis por observador e iteración, resultaron 35 raíces canónicas para explicar la totalidad de la varianza, alcanzando el 75.846% en las tres primeras raíces canónicas (ver Suplementario 1, STabla 6). Las matrices de Mahalanobis resultó en tu mayoría significativa ($p$=<0.0001) con solo cinco

**Tabla 6**. Matriz de distancias de Mahalanobis entre las observaciones resultante del análisis de variación canónica, en color se señalan aquellas distancias de mayor escalar en el estudio. Abajo se indica el valor promedio, mínimo y máximos de distancias.

| | Obs1 | Obs2 | Obs3 | Obs4 | Obs5 | Obs6 | Obs7 | Obs8 |
|---|---|---|---|---|---|---|---|---|
| Obs2 | 4.5255 | | | | | | | |
| Obs3 | 2.5493 | 2.9368 | | | | | | |
| Obs4 | 3.5053 | 2.7718 | 2.3211 | | | | | |
| Obs5 | 3.5223 | 2.394 | 2.187 | 2.5915 | | | | |
| Obs6 | 2.2812 | 4.7541 | 2.6007 | 3.7283 | 4.043 | | | |
| Obs7 | 2.7301 | 2.5566 | 1.3807 | 2.2141 | 1.8839 | 3.172 | | |
| Obs8 | 3.7625 | 2.2771 | 2.4589 | 2.1251 | 1.8906 | 4.1601 | 2.1371 | |
| Obs9 | 3.8805 | 2.3764 | 2.794 | 2.7734 | 1.9414 | 4.1825 | 2.5249 | 2.2794 |

Promedio= 2.83, mínimo= 1.38, máximo=4.75

casos de no significación (ver Suplementario 1, STabla 7), con un valor medio de 3.39 (mínimo=0.74, máximo=7.21). En este caso una gran cantidad de comparaciones realizadas por distancias de Procrustes no resultaron significativas (ver Suplementario 1, STabla 8), lo que da cuenta de las inconsistencias entre observadores.

Finalmente, el dendograma resultante para los observadores resultó en una agrupación del observador 3 y 7 con 87% de clasificaciones correctas en el remuestreo (Figura 5). Los íconos empleados en el dendograma (ramas finales) indican la especialidad y el grado de estudios, desde pasante hasta doctorado. Como se puede ver, si bien existe agrupación por grado académico, este no es sólido bajo los criterios de distancia y porcentaje de permutación. El segundo dendograma por observador e iteración (ver Suplementario 1, SFigura 1) resultó en tres agrupaciones con cuatro de las iteraciones realizadas por observador (Observadores 1, 2 y 8). Además, resultó en tres agrupaciones con tres de las cuatro iteraciones realizadas (Observadores 4, 5 y 9). También, resultó en dos agrupaciones de dos de las cuatro iteraciones realizadas (Observadores 3, 6). El observador 7 no logró ninguna agrupación entre sus iteraciones.

## Discusión

El objetivo del presente trabajo fue explorar mediante un diseño experimental anidado, el error de medición en un protocolo de *landmarks* dentales analizando su fluctuación en función del tiempo (iteraciones) para verificar si éste, se comportaba como el dogma predice: disminución del error con la experiencia del observador. Si bien existen muchas formas en las cuales realizar la caracterización del error de medición, aquí se aplicó modelos para la morfometría geométrica y algunas técnicas estadísticas bivariadas y multivariadas.

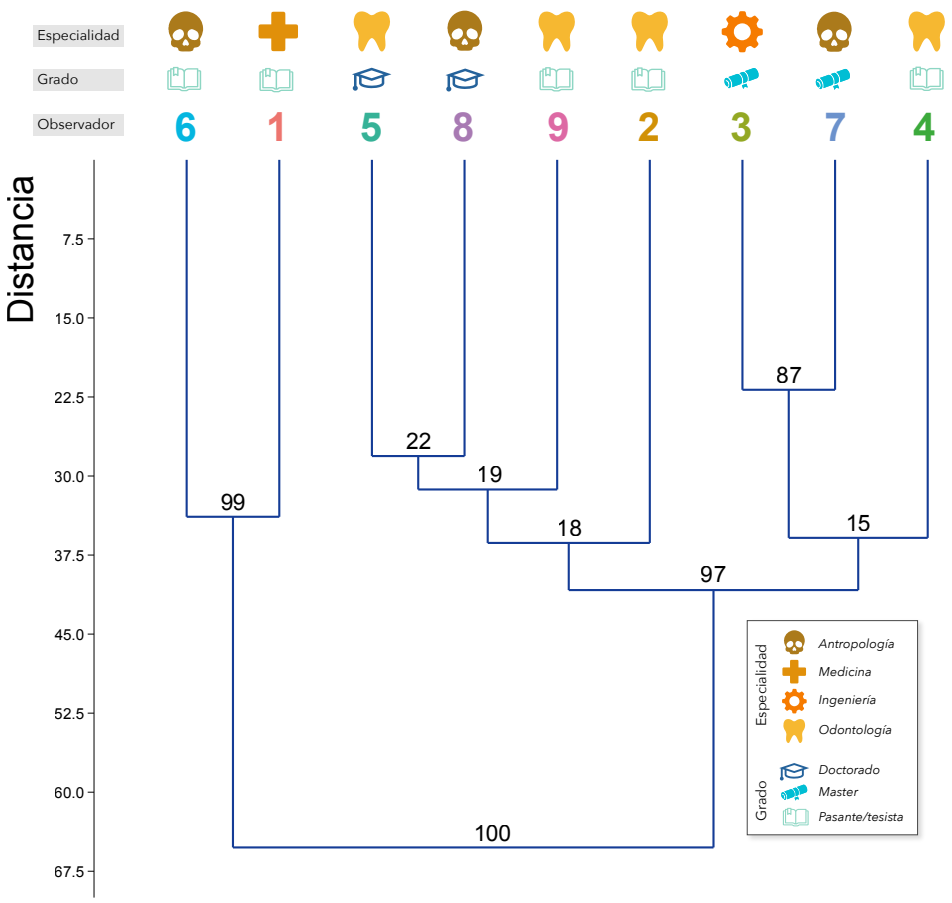


**Figura 5**. Dendograma UPGMA de los observadores en el análisis, se grafica la distancia y el porcentaje de agrupación resultante para 10,000 permutaciones. Se ha señalado la especialidad, el grado académico y el observador en la parte superior.

## *Sobre los outliers*

Como primer punto de cualquier protocolo de identificación humana, ya sea para la caracterización del error o para un estudio morfológico particular, la evaluación de los individuos fuera del rango de variación normal (*outliers*) es una exigencia protocolar básica, es por así decirlo, un control de calidad de los datos morfométricos clásicos y geométricos. En un análisis morfológico, la cuantificación precisa de la forma puede verse obstaculizada por varios factores que introducen variaciones de forma sin sentido en un conjunto de datos, por ejemplo: la distorsión de la muestra, la alteración en la preparación, y la digitalización descuidada de los puntos de referencia (39). Para Zimek y Filzmoser, (40) determinar si una observación es un valor atípico es un ejercicio subjetivo. En la morfometría clásica la inspección de outliers no era una práctica sistemática, y a lo sumo se realizaba por medio de la inspección de forma univariada por medio de gráficos boxplots. En la morfometría geométrica la identificación de *outliers* es una práctica de importancia prioritaria, un método subjetivo para detectar valores atípicos es la inspección de gráficos de PCA o

dendrogramas buscando los datos muy separados de la dispersión principal de muestras (39). Algunos programas de MG, incluyen una rutina para la identificación de valores atípicos basada en el ajuste de una distribución normal multivariante y la distancia de Procrustes para cada muestra desde la forma de consenso o promedio, ello en el programa MorphoJ (33). Aquí se preparó un intervalo para el 99% sobre los datos de distancia a la forma media obteniendo 51 casos. Es importante citar que a ninguno de los observadores se les permitió corregir los datos de sus puntos para evaluar de forma real, la cantidad de errores que se presentaron. Las frecuencias de *outliers* fueron significativas entre disciplinas (medicina=1, ingeniería=1, odontología=4, antropología biológica=3) y especialidad (pasante=5, master=2, doctorado=2). Si bien la especialización parece que es importante, es importante acotar que la práctica realizando fenotipado es realmente lo que da la precisión a través del tiempo. En el estudio todos los observadores tuvieron *outliers* relacionados a errores en el posicionamiento de *landmarks*, incluso los participantes con más grado académico, entonces dicho factor no exime a un profesional de caracterizar su error.

Finalmente, existen metodologías multivariadas para la detección de valores atípicos basados en distancias de Mahalanobis (41) lo que también puede ser aplicado a datos como los aquí tratados. Existen también aquellos que basan su identificación en la distancia de Mahalanobis de los primeros tres componentes principales robustos por medio del paquete *mvoutliers* en *R* (42,43). La aplicación de diseños experimentales basados en estadística univariada (1 ó 2 variables) suele ser menos sensibles al EM que las comparaciones estadísticas multivariadas (i.e. más de 3 variables; ver 44). Luego entonces la caracterización de los valores atípicos no debería ser un problema para con el protocolo de error de medición en morfometría geométrica.

### *Sobre el análisis de la varianza entre los morfoespacios*

En morfología, siempre se ha considerado que el nivel de experiencia que presenta el observador determina la cantidad de error, porque el grado de experiencia se encuentra inversamente relacionado con la magnitud del EM esperado (45), esto es el dogma del error de medición, un criterio subjetivo que se ha fijado en la comunidad sin estar basada en evidencia. Para este dogma la experiencia del observador determina la confiabilidad de los datos, disminuyendo en función del tiempo. Según los resultados se constató que dicho fenómeno sucedió en un solo caso de los nueve observadores incluidos en el análisis, donde se observó una disminución gradual de la varianza entre sus digitalizaciones (ver Figura 4, Tabla 1). Es importante señalar que, aunque dicho observador (Observador 5), obtuvo el patrón esperado de disminución de varianza en función de las iteraciones (tiempo), no tuvo los valores más bajos de varianza, por el contrario, su medición de menor varianza no fue cercana a la media de varianza entre observadores (Figura 4). En general, los

observadores tuvieron un patrón aleatorio, con fuertes fluctuaciones entre sus iteraciones, salvo para el observador 8 que su varianza es la mínima de todo el diseño realizado. Es así como, según los resultados, no se soporta el dogma del EM y que, por tanto, siempre, independiente del expertise del observador la caracterización del error debe ser la norma. Se ha demostrado que aunque puede haber una tendencia a la reducción del EM, esto solo ocurre después de una curva de aprendizaje prolongada, midiendo más muestras de las que se utilizan a menudo en los estudios morfométricos (15). Además, reconocer la importancia de que el error intra e interobservador siempre está presente en los estudios de morfometría, incluso en las situaciones más favorables (12,44,46–48).

Igual, se ha reportado que el EM aumenta cuando el rasgo a relevar no presenta un amplio rango de variación fenotípica en su expresión (15,17), empero en el caso de los resultados, los valores de varianza fluctúan mucho entre observadores lo que hace pensar en los efectos de error intra e interobservador. Es así como los datos de este diseño no permiten secundar la afirmación sobre la baja varianza fenotípica y el aumento del error.

### *Sobre el cálculo del error intraobservador*

Existe suma evidencia de que el error intra e inter observador puede modificar completamente los resultados obtenidos (13,49), por ejemplo de los análisis multivariados (44). En general, el error intra observador puede ser despreciado por medio de un buen diseño experimental y el entrenamiento de los observadores (6,14,15). Los resultados demuestran que, si bien los observadores pudieron ser consistentes entre sus iteraciones, algunos puntos siempre resultan problemáticos y el estudio del error en cada estructura particular debe explorar dichas tendencias, caracterizarlas y resolverlas con diseños experimentales, entrenamientos específicos y control de datos. En morfometría geométrica, se ha reportado que algunos puntos presentan mejor consistencia que otros para el error intraobservador, por ejemplo, Menéndez (6), encontró que los *landmarks* nasion, ectoconquio, temporo esfeno-parietal, asterion, y temporo-mandibular presentaron la mayor cantidad de error intraobservador. Los resultados aquí presentados solo identificaron una diferencia entre las observaciones en el *landmark* 10 del observador 3, con una congruencia entre el resto de las digitalizaciones. Corner et al. (50) establece que para reducir el error intraobservador es recomendable que el observador esté familiarizado con el equipo y que se esté entrenado en la localización y variabilidad del *landmark* entre individuos. En este caso incluso para observadores del área no biológica (ingeniería), pero con entrenamiento en biometría, resultó en el único dato con diferencias entre iteraciones.

*Sobre las diferencias interobservadores*

En antropometría se han evidenciado los efectos del error interobservador sobre el resultado de diferenciación de poblaciones (51) y en general controversias sobre las diferencias encontradas entre observadores, algunas positivas (e.g 50,52 y otras negativas 44,53,54). De manera general, los estudios han caracterizado mayor cantidad de error entre observadores, en el caso de los resultados secunda esta idea, los valores de RMSE obtenidos son del doble que en el caso del error intraobservador.

Se ha constatado que la principal fuente de EM reside en la localización incorrecta de puntos y medidas (55), en este caso y con los datos de los *outliers* podemos precisar que la fuente de error más común aquí fue la posición en la que se colocaron los *landmarks* 1, 10, 11 y 12. El *landmark* 1 es tipo II, y el 10-12 tipo I, se ha afirmado que los *landmarks* de tipo II, entendidos como puntos anatómicos cuya supuesta homología de un espécimen a otro es respaldada únicamente por la geometría y no por evidencia anatómica (e.g. glabela, eurion; (35)) son problemáticos y están asociados a magnitudes altas de EM (56). En los datos aquí analizados solo un punto es tipo II (*landmark* 1) pero este puede estar asociado a la ausencia del segundo o tercer molar. Por otro lado, para el caso de los *landmarks* 10-11 se puede argumentar que el principal efecto es la transposición. Esto es esperable, ya que existen reportes que describen que cuando se registran medidas lineales a partir de dos puntos, las distancias más cortas o caracteres más pequeños suelen tener mayor EM que las que son más largas (10,15,57), y este podría ser una de las explicaciones de la confusión en las digitalizaciones de puntos.

También, es importante destacar que en su mayoría el error interobservador quedó por arriba del 5%, lo que aunado a los valores de varianza y presencia de *outliers*, nos hace inclinarnos por la precaución sobre el error interobservador (44,53,54). El análisis de variación canónica permite complementar al estudio de error interobservador. Aquí, como en otros experimentos similares se intentó ver el efecto del observador e iteración en la diferenciación de individuos. Los resultados indicaron que la mayoría de observadores e iteraciones fueron estadísticamente diferentes, lo que refuerza la idea de que incluso con el mismo entrenamiento y experiencia las diferencias entre la observación es importante entre observadores como otros autores también han destacado (13,49,54). Algunos autores sugieren realizar un examen previo donde se identifique visualmente los puntos y medidas a registrar para reducir sustancialmente el EM (52,58,59), en el caso aquí presentado todos los observadores fueron entrenados previamente para realizar el experimento, de forma conjunta en dos sesiones, por lo que se considera que hubo una estandarización previa.

Finalmente, el estudio de cluster o conglomeradas sobre las distancias de Mahalanobis ha permitido visualizar la relación entre la proximidad de las distancias entre observadores e iteraciones realizadas. Solo en tres observadores se mostró consistencia entre sus iteraciones, uno de ellos también con la menor

varianza entre toda la muestra, pero dos más sin esta característica. Luego, entonces, este resultado secunda la visión sobre la precaución para con el error interobservador.

### *Comentarios finales*

En estudios morfológicos, se ha evidenciado que la magnitud del EM se incrementa con el número de variables registradas (17), para este caso el protocolo de puntos es similar a lo reportado en estudios previos (60,61) por que lo que este factor parece, no fue determinante en el presente estudio. De igual forma, se ha reportado una relación proporcionalmente inversa entre la longitud de pasos involucrados en la obtención de datos y la cantidad de EM que se genera (17,52), para el caso del diseño aquí planteado los pasos son estándar para cualquier trabajo morfogeométrico, por ende, se considera equiparable entre otros trabajos. También se ha generalizado la recomendación de minimizar el intervalo temporal en la recogida de datos, ya que el EM aumenta junto con el intervalo temporal que transcurre al registrar datos de distintos especímenes en variables discretas (62), el estudio consideró intervalos de una semana entre mediciones por lo cuál se considera como independiente cada una de las iteraciones realizadas.

Por ello, como conclusión se recomienda: 1) realizar mediciones por lo menos en dos sesiones independientes, 2) realizar estimaciones de error en la toma métrica para reportar el porcentaje de error en la medición (lo que es independiente del error del modelo), 3) emplear métodos de determinación repetibilidad y reproducibilidad multivariados, 4) realizar controles de datos (e.g. individuos fuera del rango de variación normal) y 5) verificar efectos estadísticos en los datos (e.g. multicolinealidad y heterocedasticidad). Todo lo anterior, permitirá que una dictaminación forense (e.g. perfil biológico, individualizante o identificación) asegure que la descripción reportada es el resultado de variación biológica y no el resultado del error de medición acumulado en la estimación realizada.

### *Agradecimientos*

El presente manuscrito es producto del Proyecto de Desarrollo científico para atender problemas nacionales de Conacyt PN-2016-2518 con el título "Proyecto NN: Protocolos de Identificación Humana, Colección Osteológica Nacional y Base de Datos de Personas No Identificadas y/o Desaparecidas". A la Dra. Elba Rosa Leyva Huerta, Directora de la Facultad de Odontología UNAM y al Dr. Luis Pablo Cruz Hervert Jefe de la División de Estudios de Posgrado e Investigación por su colaboración en la conformación de la Colección Odontológica Nacional. A los dos revisores del manuscrito que contribuyeron enormemente a su mejora.

# Bibliografía


1. MacLeod N. Geometric morphometrics and geological shape-classification systems. Earth-Sci Rev. 2002;59(1–4):27–47.
2. Von Cramon-Taubadel N, Frazier BC, Lahr MM. The problem of assessing landmark error in geometric morphometrics: Theory, methods, and modifications. Am J Phys Anthropol. 2007 Sep;134(1):24–35.
3. Merilä J, Björklund M. Fluctuating Asymmetry and Measurement Error. Syst Biol. 1995 Mar 1;44(1):97–101.
4. Llamosa L, Contreras L, Arbelaez M. Estudio de repetibilidad y reproducibilidad utilizando el método de promedios y rangos para el aseguramiento de la calidad de los resultados de calibración de acuerdo con la norma técnica NTC-ISO/IEC 17025. Sci Tech. 2007;XIII(35):455–60.
5. IUPAC. Compendium of chemical terminology. Oxford, UK: Blackwell Scientific; 1997. http://goldbook.iupac.org/.
6. Menéndez LP. Comparing Methods to Assess Intraobserver Measurement Error of 3D Craniofacial Landmarks Using Geometric Morphometrics Through a Digitizer Arm. J Forensic Sci. 2017;62(3):741–6.
7. Byrnes JF, Kenyhercz MW, Berg GE. Examining Interobserver Reliability of Metric and Morphoscopic Characteristics of the Mandible. J Forensic Sci. 2017 Jul 1;62(4):981–5.
8. Richtsmeier JT, Paik CH, Elfert PC, Cole TM, Dahlman HR. Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans. Cleft Palate Craniofac J. 1995 May;32(3):217–27.
9. Wolak ME, Fairbairn DJ, Paulsen YR. Guidelines for estimating repeatability. Methods Ecol Evol. 2012;3(1):129–37.
10. Ross AH, Williams S. Testing repeatability and error of coordinate landmark data acquired from crania. J Forensic Sci. 2008;53(4):782–5.
11. Greene DL. Fluctuating dental asymmetry and measurement error. Am J Phys Anthropol. 1984;65(3):283–9.
12. Utermohle CJ, Zegura SL. Intra- and interobserver error in craniometry: A cautionary tale. Am J Phys Anthropol. 1982 Mar 1;57(3):303–10.
13. Pérez-Pérez A, Alesan A, Roca L. Measurement error: Inter-and Intraobserver Variability. An Empiric Study. Int J Anthropol. 1990 Jun;5(2):129–35.
14. Pérez S, González P, Bernal V, Del Papa M, Barreiro A, Negro C, et al. El error de observación y su influencia en los análisis morfológicos de restos óseos humanos: Datos de variación continua. Rev Argentina Antropol Biológica. 2004;6(1):61–75.
15. Yezerinac SM, Lougheed SC, Handford P. Measurement Error and Morphometric Studies: Statistical Power and Observer Experience. Syst Biol. 1992;41(4):471–82.
16. Zelditch M, Swiderski D, Sheets H, Fink W. Geometric morphometric for biologists: a primer. London: Academic Press; 2004. 488 p.
17. Arnqvist G, Martensson T. Measurement error in geometric morphometrics: empirical strategies to asses and reduce its impact on measures of shape. Acta Zool Acad Sci Hungaricae. 1998;44(1–2):73–96.
18. Taylor B, Kuyatt C. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. NIST Technical Note 1297-DIANE Publishing; 2009. 20 p.
19. Guyatt GH. Evidence-based medicine. ACP J Club. 1991;114(A16).
20. Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. Vol. 329, British Medical Journal. BMJ Publishing Group; 2004. p. 990–1.
21. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023).
22. Sackett DL. Evidence-based medicine. Semin Perinatol. 1997 Feb;21(1):3–5.
23. NAS. Strengthening Forensic Science in the United States: A Path Forward. 2009.
24. Arnqvist G. Measurement error in geometric morphometrics Empirical strategies to assess and reduce its impact on measures of shape.pdf. Acta Zool Acad Sci Hungaricae. 1998;44:73–96.
25. Rabinovich S. Measurement Errors: theory and practice. New York: Amer Inst of Physics; 1994. 284 p.
26. Rohlf FJ. Bias and error in estimates of mean shape in geometric morphometrics. J Hum Evol. 2003;44(6):665–83.
27. Muñoz-Muñoz F, Perpiñán D. Measurement Error in Morphometric Studies: Comparison between Manual and Computerized Methods. Ann Zool Fennici. 2010;47(1):46–56.
28. Muñoz-Muñoz F, Sans-Fuentes MA, López-Fuster MJ, Ventura J. Evolutionary modularity of the mouse mandible: dissecting the effect of chromosomal reorganizations and isolation by distance in a Robertsonian system of Mus musculus domesticus. J Evol Biol. 2011;24(8):1763–76.
29. Singleton M. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). J Hum Evol. 2002;42(5), 547.

30. Neale W, Hessel D, Terpstra T. Photogrammetric measurement error associated with lens distortion. SAE Tech Pap. 2011;01–0286:1–54.
31. Kieser JA, Bernal V, Neil Waddell J, Raju S. The Uniqueness of the Human Anterior Dentition: A Geometric Morphometric Analysis. J Forensic Sci. 2007 May 1;52(3):671–7.
32. Rohlf FJ. The tps series of software. Hystrix. 2015 Jun 12;26(1):1–4.
33. Klingenberg CP. MorphoJ: an integrated software package for geometric morphometrics. Mol Ecol Resour. 2011;11(2):353–7.
34. Rohlf FJ, Slice DE. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool. 1990;39(1):40–59.
35. Bookstein FL. Morphometric tools for landmark data. Cambridge, Massachusetts,USA: Cambridge University Press; 1991. 435 p.
36. Goodall C. Procrustes Methods in the Statistical Analysis of Shape. J R Stat Soc Ser B. 1991;53(2):285–339.
37. Singleton M. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). J Hum Evol. 2002;42(5):547–78.
38. Sokal RR, Rohlf FJ. Biometry: the principles and practice of statistics in biological research. San Francisco: W. H. Freeman; 1995. 880 p.
39. Renner MAM, Brown EA, Wardle GM, Renner MAM, Brown EA, Wardle GM. Averaging v. outlier removal. Decrypting variance among cryptic Lejeunea species (Lejeuneaceae: Jungermanniopsida) using geometric morphometrics. Aust Syst Bot. 2013 Apr 12;26(1):13–30.
40. Zimek A, Filzmoser P. There and back again: Outlier detection between statistical reasoning and data mining algorithms. Wiley Interdiscip Rev Data Min Knowl Discov. 2018 Nov 1;8(6):e1280.
41. Viscosi V, Fortini P, Slice DE, Loy A, Blasi C. Geometric morphometric analyses of leaf variation in four oak species of the subgenus Quercus (Fagaceae). http://dx.doi.org/101080/11263500902775277. 2009 Nov;143(3):575–87.
42. Filzmoser P. Identification of Multivariate Outliers: A Performance Study. Austrian J Stat. 2005 Apr 3;34(2):127–138–127–138.
43. Filzmoser P, Gregorich M. Multivariate Outlier Detection in Applied Data Analysis: Global, Local, Compositional and Cellwise Outliers. Math Geosci 2020 528. 2020 Apr 2;52(8):1049–66.
44. Jamison PL, Zegura SL. A univariate and multivariate examination of measurement error in anthropometry. Am J Phys Anthropol. 1974 Mar 1;40(2):197–203.
45. Gordon CC, Bradtmiller B. Interobserver error in a large scale anthropometric survey. Am J Hum Biol. 1992;4(2):253–63.
46. Utermohle CJ, Zegura SL, Heathcote GM. Multiple observers, humidity, and choice of precision statistics: Factors influencing craniometric data quality. Am J Phys Anthropol. 1983 May 1;61(1):85–95.
47. Gaito J, Gifford EC. Components of Variance in Anthropometry. Hum Biol. 1958;30(2):120.
48. Kemper HCG, Pieters JJL. Comparative study of anthropometric measurements of the same subjects in two different institutes. Am J Phys Anthropol. 1974 May 1;40(3):341–3.
49. Perez SI, Gonzalez PN, Bernal V, Del Papa M, Barreiro A, Negro C, et al. El error de observación y su influencia en los Análisis morfológicos de restos óseos humanos. Datos de variación continua. Rev Argentina Antropol Biológica. 2004;6(1):61–75.
50. Corner B, Lele S, Richtsmeier R. Measuring precision of three-dimensional landmark data. J Quant Anthr. 1992;3:347–59.
51. Bennett KA, Osborne RH. Interobserver measurement reliability in anthropometry. Hum Biol. 1986;58(5):751–9.
52. Sholts SB, Wärmländer SKTS, Flores LM, Miller KWP, Walker PL. Variation in the measurement of cranial volume and surface area using 3d laser scanning technology. J Forensic Sci. 2010 Jul;55(4):871–6.
53. Robinson C, Terhune CE. Error in geometric morphometric data collection: Combining data from multiple sources. Am J Phys Anthropol. 2017 Sep 1;164(1):62–75.
54. Shearer BM, Cooke SB, Halenar LB, Reber SL, Plummer JE, Delson E, et al. Evaluating causes of error in landmark-based data collection using scanners. PLoS One. 2017 Nov 1;12(11):e0187452.
55. Ossenberg N. Within and between race distance in population studies based on discrete traits of the human skull. Am J Phys Anthr. 1979;45:701–16.
56. Algee-Hewitt BFB, Wheat AD. The reality of virtual anthropology: Comparing digitizer and laser scan data collection methods for the quantitative assessment of the cranium. Am J Phys Anthropol. 2016 May 1;160(1):148–55.
57. Kouchi M, Mochimaru M, Tsuzuki K, Yokoi T. Interobserver errors in anthropometry. J Hum Ergol (Tokyo). 1999 Dec 15;28(1–2):15–24.
58. Fourie Z, Damstra J, Gerrits PO, Ren Y. Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. Forensic Sci Int. 2011 Apr 15;207(1–3):127–34.
59. Sholts SB, Flores L, Walker PL, Wärmländer SKTS. Comparison of coordinate measurement precision of different

landmark types on human crania using a 3D laser scanner and a 3D digitiser: Implications for applications of digital morphometrics. Int J Osteoarchaeol. 2011 Sep 1;21(5):535–43.

60. Jasso-Cuéllar J, Gil-Chavarría I, Quinto-Sánchez M. Anterior dental arch shape and human identification: Kieser et al. method applied to 2D-3D dental models in Mexican population. Forensic Sci Int Reports. 2020 Nov 1;2:100161.
61. Sheets H., Bush PJ, Bush MA. Patterns of Variation and Match Rates of the Anterior Biting Dentition: Characteristics of a Database of 3D-Scanned Dentitions. J Forensic Sci. 2013 Jan;58(1):60–8.
62. Molto JE. The assessment and meaning of intraobserver error in population studies based on discontinuous cranial traits. Am J Phys Anthropol. 1979;51(3):333–44.