



## Sex Determination in Forensic Dentistry Using Supervised Classification Techniques

Ram{\'n} \'lvarez-Vaz, Carlos Sassi

### ► To cite this version:

Ram{\'n} \'lvarez-Vaz, Carlos Sassi. Sex Determination in Forensic Dentistry Using Supervised Classification Techniques. 2020. hal-02486444

**HAL Id: hal-02486444**

**<https://hal.science/hal-02486444>**

Submitted on 21 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## DETERMINACIÓN DEL SEXO MEDIANTE TÉCNICAS DE CLASIFICACIÓN SUPERVISADA<sup>a</sup>

## SEX DETERMINATION IN FORENSIC DENTISTRY USING SUPERVISED CLASSIFICATION TECHNIQUES

RAMÓN ÁLVAREZ-VAZ<sup>b</sup> \*, CARLOS SASSI<sup>c</sup>

Recibido 25-06-2019, aceptado 01-10-2019, versión final 12-11-2019.

Artículo Investigación

**RESUMEN:** Los dientes, componentes esenciales del sistema estomatognático, constituyen un excelente material para investigaciones paleontológicas, arqueológicas, antropológicas, genéticas y forenses, merced a su fácil accesibilidad, examen, registro y estructura casi indestructible. La información referente a su tamaño y características ha resultado de suma utilidad en la determinación del sexo y en el propio proceso de identificación humana. Este estudio buscó verificar si el índice canino maxilar, era un instrumento confiable y válido para la determinación del sexo, en una muestra uruguaya. Se trató de un análisis odontométrico de 481 modelos superiores de yeso (243 de sujetos del sexo masculino y 238 del femenino), de pacientes con edades comprendidas entre 18 y 60 años, asistidos en una clínica de ortodoncia de la ciudad de Montevideo, Uruguay, consistente en las mediciones del diámetro mesiodistal de los caninos y la distancia intercanina superiores, y el cálculo del correspondiente índice canino maxilar. Se compararon tres técnicas de clasificación del sexo, el índice canino maxilar y dos métodos de clasificación supervisada (árbol de clasificación y árbol particionado recursivo), buscando verificar qué método era el más adecuado para la determinación del sexo, en una muestra uruguaya. Los modelos mostraron una performance de 52 %, 77 % y 69 % respectivamente. Se concluyó que los modelos estadísticos de clasificación supervisada, permitieron realizar una determinación del sexo de manera más fidedigna que el método de discriminación convencional.

**PALABRAS CLAVE:** Dimorfismo sexual; odontometría; índice canino; clasificación supervisada.

**ABSTRACT:** Teeth, essential components of stomatognathic system, constitute an excellent material for paleontological, archaeological, anthropological, genetic and forensic researches by virtue of their easy accessibility, examination, registration and almost indestructible structure. Information regarding their size and characteristics has been extremely useful in determining sex and in the process of human identification itself. This survey was an odontometric analysis of 481 upper dental stone casts (243 from males and 238 from females), of patients aged between 18 and 60 years, assisted in one orthodontic clinic of Montevideo, Uruguay, consisting of measurements of upper mesiodistal

<sup>a</sup>Álvarez-Vaz, R. & Sassi, C. (2020). Índice Canino Maxilar: Determinación del sexo mediante técnicas de clasificación supervisada. *Rev. Fac. Cienc.*, 9 (1), 6–24. DOI: <https://doi.org/10.15446/rev.fac.cienc.v9n1.56684>

<sup>b</sup>M. Sc. en Epidemiología. Profesor Agregado, Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.

\* Autor para correspondencia: [ramon@iesta.edu.uy](mailto:ramon@iesta.edu.uy)

<sup>c</sup>Ph. D. en Biología Bucodental. Facultad de Odontología de Piracicaba, Universidad de Campinas, San Pablo, Brasil.

diameter of canines and intercanine distance. Three sex classification techniques were compared, the maxillary canine index and two supervised classification methods (classification tree and recursive partitioned tree), seeking to verify which method was the most suitable for the determination of sex, in a Uruguayan sample. The models showed a performance of 52%, 77% and 69% respectively. It was concluded that the statistical models of supervised classification allowed a determination of the sex in a more reliable way than the discrimination method.

**KEYWORDS:** Sex dimorphism; odontometry; canine index; supervised classification.

## 1. INTRODUCCIÓN

La identificación de restos humanos, uno de los mayores retos de los equipos forenses, deberá asentarse en la acción sistematizada de las múltiples especialidades involucradas en este dinámico e intrincado proceso técnico-científico (medicina y odontología legal, antropología, genética, química, entomología, geología, botánica, balística, etc.), con el designio de establecer las peculiaridades privativas de un individuo, (Clark, 1994; Campos-Neto & Vanrell, 2014). Para ello, será menester llevar a cabo una adecuada reconstitución de su perfil biológico, valiéndose de sus cuatro principales componentes: edad, sexo, estatura y ancestralidad, como lo tratan Prabhu & Acharya (2009); Pereira *et al.* (2010); Nahidh *et al.* (2013).

Los órganos dentales se erigen como una magnífica materia prima para la materialización de procedimientos comparativos y reconstructivos post mortem, en vista de su asombrosa dureza y resistencia a los más variados agentes físicos, químicos y biológicos (Harvey, 1975; Acharya & Mainali, 2009; Acharya *et al.*, 2011), y sus particularidades anatómicas y volumétricas, que pueden exteriorizar cierto grado de dimorfismo sexual (Harvey, 1975; Rai & Anand, 2007; Prabhu & Acharya, 2009), siendo pasibles de un minucioso análisis odontométrico (Rao *et al.*, 1989; Rai & Anand, 2007; Zorba *et al.*, 2011).

El carácter dimórfico de la dentición humana ha sido ratificado por numerosos investigadores en dispares poblaciones universales, a lo largo del tiempo, echando mano a métodos fundamentados en dimensiones dentales, realizadas sobre modelos de yeso. Las más comúnmente manipuladas estriban en los diámetros mesiodistal, vestibulopalatino, mesiovestibular-distopalatino y distovestibular-mesioopalatino, altura gingivoincisal, y distancia intercanina, *per se* o como parte de índices (Rao *et al.*, 1989; Rai & Anand, 2007; Prabhu & Acharya, 2009; Acharya & Mainali, 2009; Pereira *et al.*, 2010; Eboh & Etetafia, 2010; Acharya *et al.*, 2011; Zorba *et al.*, 2011; Picapedra *et al.*, 2012; Sassi *et al.*, 2012; Nahidh *et al.*, 2013; Gargano *et al.*, 2014).

Los caninos permanentes son señalados como elementos cardinales para la práctica forense (Rao *et al.*, 1989; Kaushal, 2003), al resultar los de mayor dimorfismo sexual (Kaushal, 2003; Rai & Anand, 2007; Srivastava, 2010) y sobrevida, incluso en circunstancias desfavorables, acaso por su extensa superficie de implantación, menor susceptibilidad a las paradenciopatías y exposición al biofilm, sarro, abrasión y/o carga oclusal (Rao *et al.*, 1989; Srivastava, 2010; Nahidh *et al.*, 2013).

La odontología legal y/o forense ha contado con la invaluable colaboración de dos ciencias en constante e impetuosa expansión, la estadística y la informática, capaces de elaborar arquetipos matemáticos adaptados a las necesidades de cada caso (Costa *et al.*, 2005).

Teniendo en cuenta lo expuesto, el presente trabajo procuró verificar si el índice canino maxilar y dos métodos de clasificación supervisada, constituían instrumentos fidedignos y válidos para la determinación del sexo, en una muestra uruguaya.

## 2. METODOLOGÍA

Esta investigación radicó en un estudio descriptivo, transversal y cuantitativo de 481 modelos superiores de yeso (243 de sujetos del sexo masculino y 238 del femenino), pertenecientes a pacientes asistidos en la clínica de ortodoncia del IUCEDDU (Instituto Universitario Centro de Estudio y Diagnóstico de las Disgnacias del Uruguay), con edades comprendidas entre 18 y 60 años, dentición permanente sana, incisivos y caninos totalmente erupcionados, sin restauraciones, aparatología ortodóncica, desgaste ni anomalías dentales, quienes expresaron, libre y espontáneamente, por medio de consentimiento informado, su pretensión y decisión de formar parte de la misma.

Es procedente hacer notar que el proyecto de esta iniciativa científica fue desarrollado, siguiendo las recomendaciones del CEP (Comitê de Ética em Pesquisa/Comité de Ética en Pesquisa o Investigación) de la Faculdade de Odontologia de Piracicaba– UNICAMP (Facultad de Odontología de Piracicaba– Universidad de Campinas), San Pablo, Brasil, y aprobado el 24 de agosto de 2009, según consta en el protocolo de resolución 022/2009.

Una vez en poder de los modelos de yeso, se procedió a efectuar algunas medidas y cálculos. El diámetro mesiodistal (DMD) se definió como la máxima distancia lineal entre las caras proximales de los caninos superiores (13 y 23), medido a nivel de los respectivos puntos de contacto, sirviéndose de un compás de punta seca de Korhaus (Dentaurum®), dispuesto perpendicularmente al eje mayor dental. La distancia intercanina (DIC), también designada ancho del arco canino, fue concebida como el segmento lineal demarcado por las puntas cuspídeas de los caninos superiores (13 y 23), midiéndose con el auxilio de un calibre digital de puntas finas, con una resolución de 0,01 mm. (150 mm. - Digimess®, San Pablo, Brasil), tal cual se observa en la Figura 1.

Para cumplir con el debido proceso de calibración del único operador, todas las medidas se consumaron en tres ocasiones diferentes, con un intervalo no menor a una semana entre ellas, sobre 25 modelos de trabajo seleccionados aleatoriamente de la muestra. El Coeficiente de Correlación Intraclass (ICC) intraoperador mostró en este caso, para la DIC, un valor de 0,976 y para el diámetro mesiodistal izquierdo (DMDi) de 0,897.

Se determinó el índice canino maxilar (ICMax) para el hemiarco derecho (ICMaxd) e izquierdo (ICMaxi), para cada individuo, las medias, desvíos estándares y los respectivos puntos de corte de aquellos, en base a los algoritmos empleados por Rao *et al.* (1989), para la mandíbula (ICM= cociente entre el DMD y la DIC) y se aplicaron otras dos técnicas de clasificación supervisada.

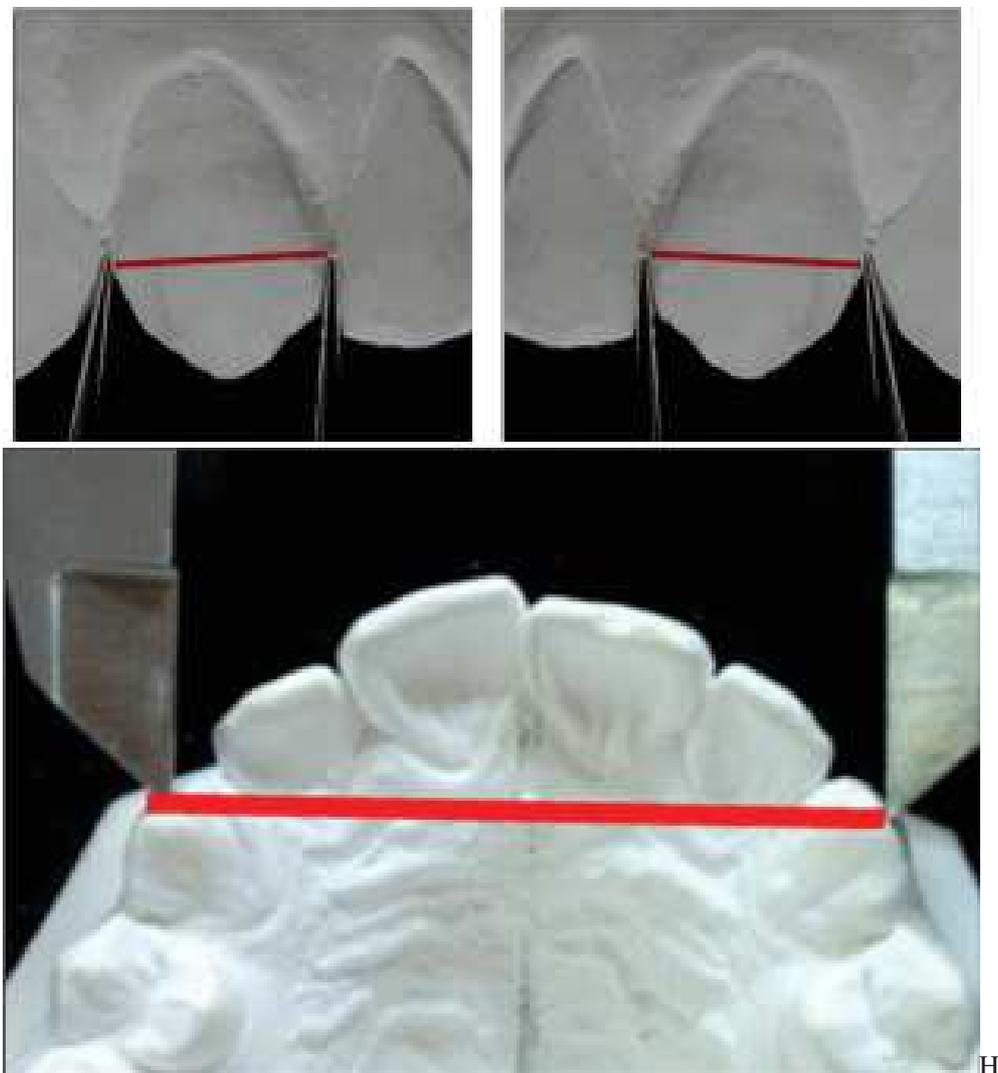


Figura 1: Medida del Ancho Mesiodistal y Distancia Intercanina. Fuente: Elaboración propia

### 3. TÉCNICAS DE CLASIFICACIÓN PROPUESTAS

En esta sección se presentan las tres técnicas que luego se compararán en la sección de aplicación, siendo la primera la que surge del artículo de Rao *et al.* (1989) y las otras dos diferentes técnicas estadísticas de clasificación supervisada.

#### 3.1. Modelo de discriminación de Rao

Los indicadores necesarios a ser relevados para el Índice de Rao aparecen en la Figura 1. Rao define el ICM como

$$\text{ICM} = \frac{\text{ancho de corona mesiodistal de caninos mandibulares}}{\text{ancho arco canino mandibular}} \quad (1)$$

a partir del cual computa el **ICM estándar**

$$\text{ICM estándar} = \frac{(IMC_1 - DE_1) + (IMC_2 + DE_2)}{2} \quad (2)$$

donde  $IMC_1$  es el Índice Promedio Mandibular Canino para hombres y el  $IMC_2$  es el Índice Promedio Mandibular Canino para mujeres.

A su vez  $DE_1$  y  $DE_2$  representan las desviaciones estándar respectivamente de hombres y mujeres.

Dado las características anatómicas de la mandíbula y maxilar superior, así como los caninos de ambos maxilares, es esperable que la aplicación al maxilar superior del método original de Rao, sea válida.

#### 3.2. Métodos CART

Este método permite la construcción de modelos basados en técnicas no paramétricas, lo que supone muchas menos restricciones de distribuciones de probabilidad en las variables consideradas, permitiendo encontrar las variables que mejor discriminan el comportamiento de una variable de respuesta o dependiente de tipo categórica; la aplicación inmediata es sobre variables que clasifican en ausencia o presencia de una patología, en diferentes niveles de patología (por ejemplo la maloclusión en escolares), (Álvarez-Vaz, 2009). La gran ventaja de estas técnicas es que prescinden de un modelo analítico explícito (como puede ser los modelos de regresión lineal múltiple, de regresión logística o análisis discriminante probabilístico), lo que las hace más fácilmente usables e interpretables por los no especialistas en estadística, (Abernathy *et al.*, 1987).

El método CART (Classification and Regression Trees) fue introducido, en 1984 por Breiman *et al.* (1984). Es una herramienta de análisis exploratorio de datos con el objetivo de encontrar reglas de clasificación y de predicción. Es uno de los métodos de aprendizaje inductivo supervisado no paramétrico más utilizado. Se

aplica en diversas disciplinas, y se destaca por su fácil interpretación y aplicabilidad.

El CART puede ser utilizado tanto en variables independientes categóricas como continuas. Si la variable de respuesta es continua, se obtienen árboles de regresión y son de clasificación, cuando la variable de respuesta es categórica. Se caracteriza por no requerir una selección a priori de variables, presencia de interacciones o transformaciones de variables y es de más fácil interpretación que los modelos de regresión tradicionales.

El método es relativamente sencillo y consiste en la construcción de un árbol de decisiones, muy similar al que se usa en las guías clínicas. Se comienza con un nodo raíz que contiene todas las observaciones, éste es dividido en subgrupos determinados por la partición de la variable elegida, generando nodos descendentes. Estos subgrupos son divididos usando la dicotomización de una segunda variable, y así sucesivamente hasta alcanzar los nodos terminales, que es cuando se logra un grupo lo más homogéneo posible, obteniéndose la mayor representación de una clase. Se consideran todas las particiones posibles, siendo cada partición jerarquizada según un criterio de calidad.

La regla de clasificación es sencilla: en cada nodo de decisión se verifica si el valor de cierta variable es mayor que cierto valor específico. Si es mayor se sigue el camino de la derecha y si es menor, el de la izquierda.

Para el proceso de partición existen un conjunto de preguntas binarias  $Q$ . Si la variable es cuantitativa la pregunta será del tipo:  $X_m \leq v$  y si la variable es cualitativa la pregunta será del tipo:  $x_m \in S$  siendo  $S$  un subconjunto de  $(b_1, b_2, \dots, b_L)$  existiendo  $L$  clases posibles para la variable  $X$ . Si la respuesta es positiva va a la izquierda y si la respuesta es negativa va a la derecha. Se generan así, nodos “hijos” que surgen de la partición del nodo anterior. El proceso de partición de los nodos culmina con la obtención del árbol máximo; luego se puede proceder a la poda, reduciendo los nodos terminales hasta llegar a un árbol óptimo.

Las particiones tienen como objetivo incrementar la homogeneidad de los subconjuntos resultantes de la misma. Existe asociada a cada partición una *medida de impureza*. Cuando la impureza es mínima las observaciones de un nodo pertenecen a una misma clase, mientras que cuando es máxima, las observaciones de un nodo pertenecen en igual proporción a las distintas clases. La medida de impureza de un nodo, es el resultado de evaluar la *función de impureza* en ese nodo, tomando como información las proporciones  $p(j/t)$ , probabilidad de que una observación que está en el nodo  $t$  pertenezca a la categoría  $j$ .

Existe una impureza global (para todo el árbol) y la impureza de cada nodo.

Se denomina  $\Phi$  a la función de impureza

$$i(t) = \Phi(p(1/t), p(2/t), \dots, p(J/t)) \quad (3)$$

con  $p(j/t) = \frac{N_j(t)}{N(t)}$ .

Como medidas de impureza pueden manejarse:

1. índice de Gini:  $\sum p(i/t)p(j/t) = i(t)$
2. Deviance:  $D(t) = -2\sum n_{jt}p(j|t)\log(p(j|t))$

La evaluación de cada sub-árbol se realiza mediante estimaciones de las tasas de clasificación incorrecta (por sustitución, por muestra de validación o por validación cruzada).

Resumiendo, las ventajas de usar CART son:

- puede ser aplicado a cualquier tipo de variables predictoras: continuas y categóricas;
- los resultados son fáciles de entender e interpretar;
- puede trabajar con datos faltantes;
- hace automáticamente la selección de variables;
- es invariante a transformaciones de las variables predictoras;
- es robusto a la presencia de *outliers*;
- es un clasificador no paramétrico, es decir, que no requiere suposiciones de distribución de probabilidades;
- toma en cuenta las interacciones que pueden existir entre las variables predictoras;
- es rápido de calcular y muy fácil de representarse visualmente para los no especialistas.

Entre las desventajas se encuentran:

- el proceso de selección de variables es sesgado hacia las variables con más valores diferentes;
- dificultad para elegir el árbol óptimo;
- requiere un gran número de datos para asegurarse que la cantidad de observaciones en los nodos terminales sea importante.

## 4. APLICACIÓN AL ESTUDIO URUGUAYO

Tabla 1: Número de maxilares evaluados por sexo según tipo de maxilar

Maxilar	Masculino	Femenino	Total
Superior	243	238	481

#### 4.1. Medidas y cálculos efectuados

Las variables que están en la tabla de datos y que se describen a continuación son las que se obtuvieron mediante el proceso de medición o que se calcularon a partir de éstas, tal como se detalló en la sección 2.

Tabla 2: Conjunto de variables utilizadas

Descripción de Variables		
Variable	Nombre	Descripción
V(1)	DMDd	diámetro mesiodistal derecho
V(2)	DMDi	diámetro mesiodistal izquierdo
V(3)	DIC	distancia intercanina
Índices calculados		
V(4)	ICMaxd	Índice Canino Maxilar derecho
V(5)	ICMaxi	Índice Canino Maxilar izquierdo

A partir de estos valores se establecen los puntos de corte siguiendo la metodología de Rao ecuación (4). Utilizando la ecuación (2) se fija un punto de corte en 0.274, de manera que si el ICM estándar $_i \leq 0.274$  el individuo *-iésimo* se clasifica como hombre y si ICM estándar $_i > 0.274$  la persona se clasifica como del sexo femenino.

Tabla 3: Índices caninos maxilares estándares

Índice Canino Maxilar derecho por sexo							
Sexo	Min	$Q_1$	Mediana	$\bar{x}$	$Q_3$	Max.	NA's
M	0.1983	0.2215	0.2323	0.2341	0.2430	0.3187	16
F	0.1392	0.2213	0.2325	0.2335	0.2433	0.3096	12
Índice Canino Maxilar izquierdo por sexo							
M	0.1988	0.2209	0.2310	0.2332	0.2423	0.3889	16
F	0.1390	0.2205	0.2313	0.2320	0.2420	0.3377	12

$$ICMDE = (0.233 + 0.0203 + 0.234 - 0.019)/2 = 0.234 \quad (4)$$

$$ICMIE = (0.232 + 0.0295 + 0.233 - 0.025)/2 = 0.2345 \quad (5)$$

Tabla 4: Tabla de clasificación para sexo usando algoritmo de Rao

Clasificación según algoritmo de Rao para Índice Canino Maxilar derecho			
	M	F	Total
	M	F	Total
Pronóstico (M)	128	122	282
Pronóstico (F)	99	104	203
Total	227	226	453

Clasificación según algoritmo de Rao para Índice Canino Maxilar izquierdo			
	M	F	Total
	M	F	Total
Pronóstico (M)	135	126	261
Pronóstico (F)	92	100	192
Total	227	226	453

Se opta por mostrar el comportamiento de los índices para el lado izquierdo, aunque para el lado derecho los resultados son prácticamente iguales. En las Figuras 2 y 3 se puede observar cual es la densidad condicional del sexo de acuerdo a dos variables que son el DMDi y el IMC variables originales y por otra parte el ICM que sigue el algoritmo de Rao. En este caso se muestra para el ICMi, como aparece en la Figura 4.

Por otra parte se muestra como es la relación bivariada de las dos medidas usadas por Rao para la elaboración del ICM.

Observando las Figuras 2 y 3 ya se puede ver que la distribución condicional del sexo al variar el ICMi de Rao tiene un comportamiento que no es monótono creciente lo que sería un indicio de que no hay una buena diferenciación por sexo en esas dos variables, mientras que la proporción de mujeres parece decrecer al aumentar el DMDi. Los mismos aspectos se pueden evaluar cuando se analiza el comportamiento multivariado de las variables y el Índice de Rao, lo que justifica el análisis que aparece en las siguientes secciones, donde solamente se usa las mismas variables utilizadas por Rao pero con otras herramientas estadísticas.

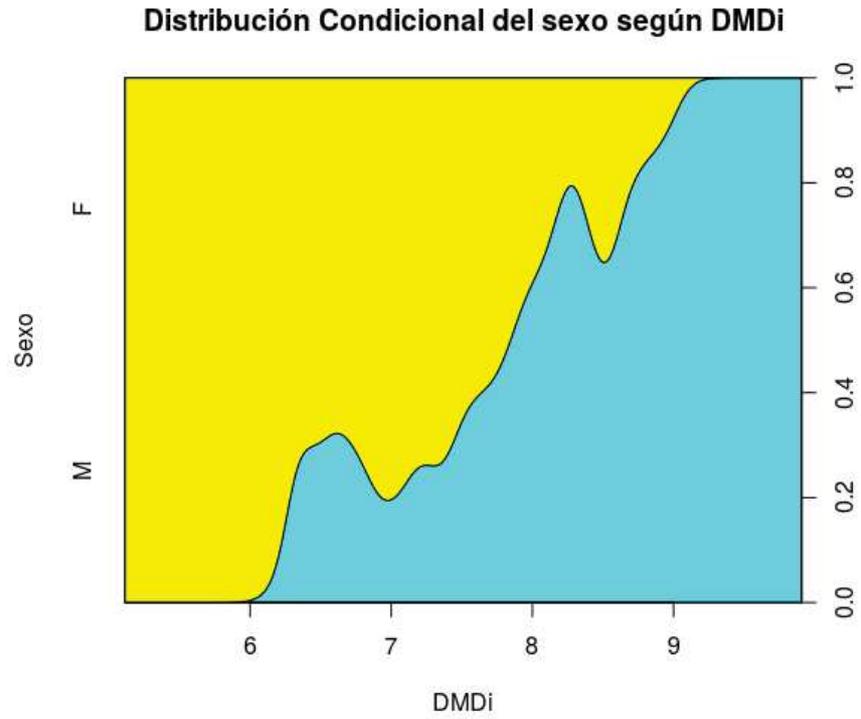


Figura 2: Proporción de M y F según DMDi para maxilar superior. Fuente: Elaboración propia.

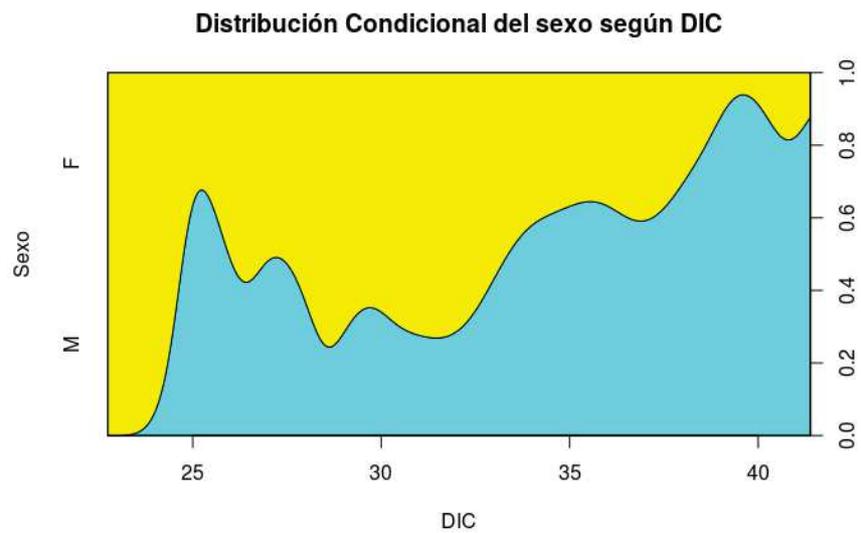


Figura 3: Proporción de M y F según DIC, para maxilar superior. Fuente: Elaboración propia.

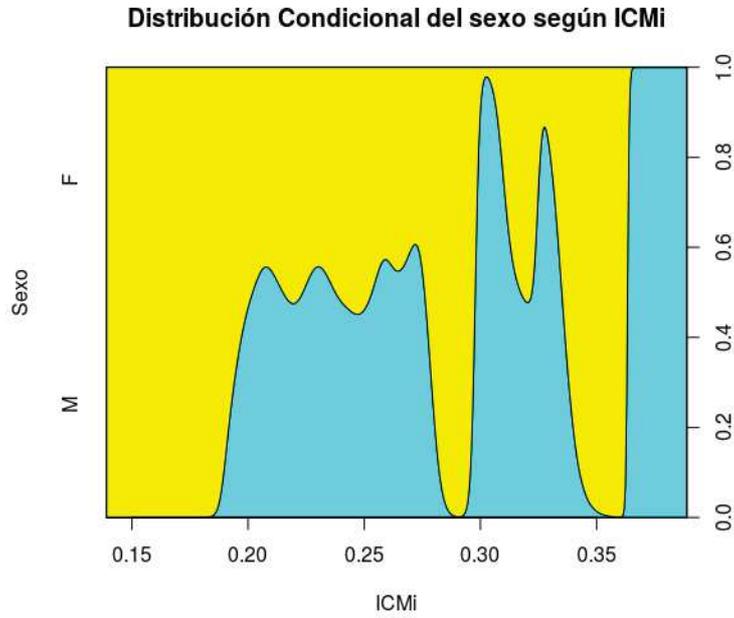


Figura 4: Proporción de M y F según ICMi, para maxilar superior. Fuente: Elaboración propia.

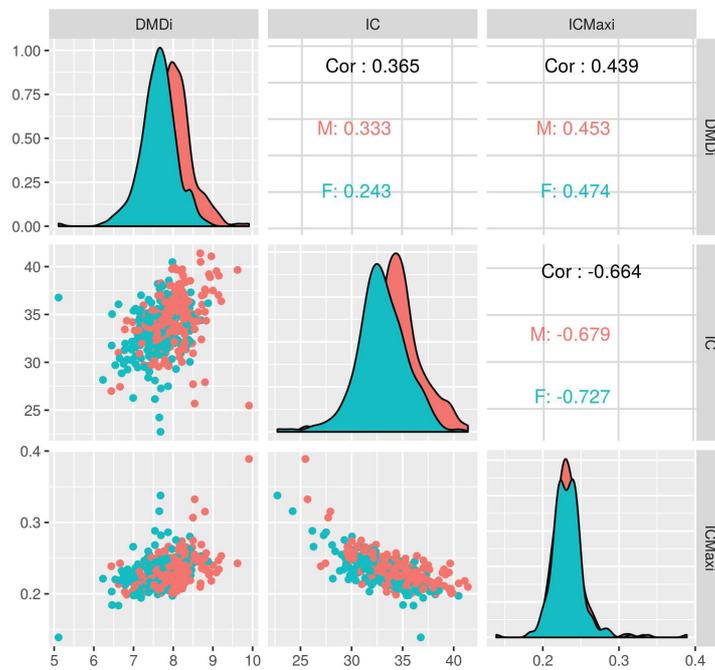


Figura 5: Relación para caninos izquierdos para maxilar superior. Fuente: Elaboración propia.

## 4.2. Performance del método CART

En este caso se estima una versión convencional **CART** donde, en base a criterios de impureza de los nodos y una función que evalúa la calidad de la partición, se elaboran los nodos y se debe decidir donde *podar* (que refiere donde cortar el árbol aislando grupos de individuos o nodos intermedios) bien diferenciados internamente y no llegar a los nodos terminales que serían todos los individuos por separado, (Therneau & Atkinson, 2018).

Por otra parte se utiliza una versión más nueva y complementaria de CART que permite hacer un particionado recursivo mediante inferencia condicional en los nodos. Es decir que a pesar de ser un modelo no paramétrico, permite en cada rama del árbol establecer si existe una diferencia significativa en cada nodo donde se establece un umbral de corte para las variables que en este caso son continuas; a la vez que construye el árbol con mayor capacidad de discriminación se está elaborando una prueba de hipótesis en forma secuencial. Este tipo de análisis se logra por medio del uso de la librería *party*, (Hothorn *et al.*, 2006).

Tabla 5: Indicadores de calidad de ajuste para 'poda' del árbol estimado

fórmula = Sexo.rec DMDi + ICM				
Root node error: 226/453 = 0.4989 n= 453				
Costo de Complejidad (CC)	nsplit	error relativo		
0.376	0	1.00		
0.020	1	0.62		
0.020	4	0.56		
0.010	8	0.51		
0.010	13	0.45		
Importancia de las variables				
DMDi	DIC	Punto de corte (PC)	Árbol	
56	44	-	Original	
64	36	0.02	Podado	

De la Tabla 5 debe considerarse que se establece un valor de  $PC = 0.02$ , para establecer donde efectuar la poda y se observa que hay un cambio en la importancia relativa de la variable DMDi con respecto a IMC. Luego de la poda se obtienen cinco nodos, como aparecen en la Figura 6, donde los nodos terminales aparecen como rectángulos mientras que los nodos intermedios aparecen como óvalos. La idea de como interpretar cada nodo es comparar la proporción de F que se logra obtener diferenciándose lo más posible de la proporción original del nodo *raíz*.

En el reporte que sigue puede verse cuál es el árbol que resulta de considerar la alternativa de inferencia condicional.

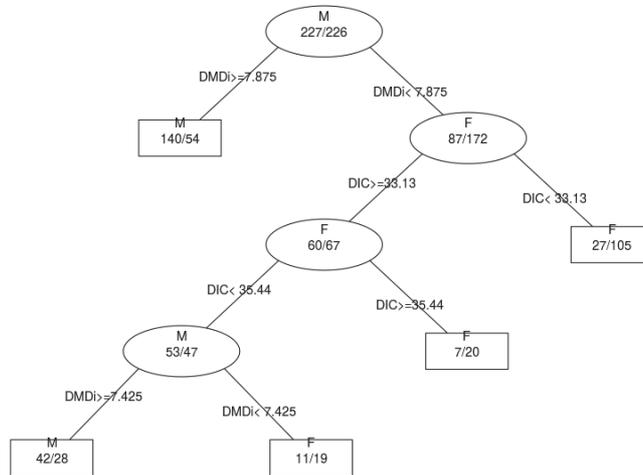


Figura 6: Árbol de Clasificación para sexo. Fuente: Elaboración propia.

Tabla 6: Árbol de Inferencia Condicional para 4 nodos terminales

Árbol de Inferencia Condicional con 4 nodos terminales

Respuesta: Sexo

Inputs: DMDi, ICM

Número de observaciones: 453

- 1)  $DMDi \leq 7.87$ ; criterion = 1, statistic = 64.382
- 2)  $DIC \leq 33.12$ ; criterion = 0.991, statistic = 8.023
- 3)\* weights = 132
- 2)  $DIC > 33.12$
- 4)\* weights = 127
- 1)  $DMDi > 7.87$
- 5)  $DIC \leq 33.71$ ; criterion = 0.969, statistic = 5.852
- 6)\* weights = 69
- 5)  $DIC > 33.71$
- 7)\* weights = 125

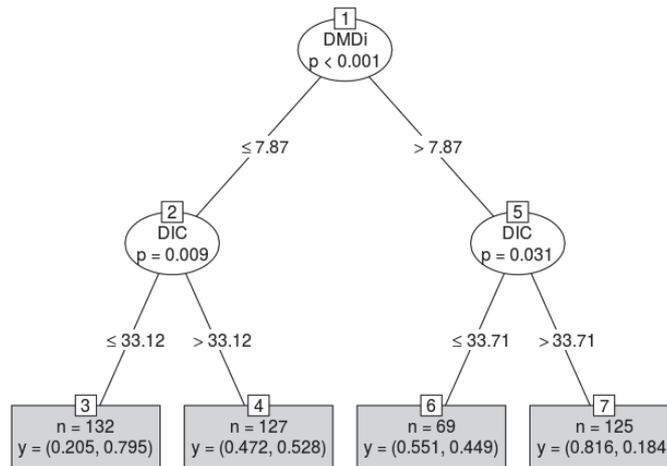


Figura 7: Árbol de Clasificación condicional para sexo. Fuente: Elaboración propia.

La interpretación para el árbol mediante particionado recursivo en cuanto a los nodos es la misma que para el caso del CART convencional.

## 5. DISCUSIÓN

Luego de haber aplicado los tres métodos de clasificación surgen tres Índices que se denominarán Índice Canino Maxilar Estándar según algoritmo de Rao, Índice Canino por Árbol de Clasificación (ICACa, ICACb), de los cuales se hace un resumen en cuanto a su performance.

Antes de analizar cada uno en particular se presentan las matrices de confusión, donde se puede ver exactamente, como funciona el pronóstico para cada categoría a clasificar. Las matrices de confusión tienen la ventaja de al estar en este caso consignadas en frecuencias absolutas permiten al lector la posibilidad de hacer un análisis en perfiles filas y en perfiles columnas (es decir, para comparar filas y /o columnas)

Los resultados encontrados si se comparan con Sassi *et al.* (2012), son similares ya que la performance en ese estudio muestra con el algoritmo de Rao aplicado en el maxilar superior una muy pobre performance pronosticando, 27 hombres clasificados correctamente (de 55), y 31 mujeres correctamente clasificadas (de 63), lo que se traduce en una tasa de éxito del 49% para ambos sexo y en forma global. En Gargano *et al.* (2014) el valor estándar del ICM fue establecido en 0.267 y el del ICMax en 0.232, por lo cual, valores menores a los referidos corresponderían al sexo femenino, al tiempo que los mayores o iguales, al masculino. El intento de validación de los índices caninos arrojó los siguientes resultados:

Tabla 7: Matrices de confusión para los tres métodos de clasificación usados.

Clasificación según diferentes algoritmos para Índice Canino Maxilar izquierdo				
Valores observados				
Algoritmo de Rao	M	F	Total	
Pronóstico (M)	<b>135</b>	126	261	
Pronóstico (F)	92	<b>100</b>	192	
Total	227	226	453	
Árbol Clasificación	M	F	Total	
Pronóstico (M)	<b>171</b>	46	217	
Pronóstico (F)	56	<b>180</b>	192	
Total	227	226	453	
Árbol particionado recursivo	M	F	Total	
Pronóstico (M)	<b>140</b>	54	194	
Pronóstico (F)	87	<b>172</b>	259	
Total	227	226	453	

Tabla 8: Performance de los diferentes Índices de clasificación

Comparación de los diferentes métodos de clasificación					
Índice	Método	Tasa de acierto Global	Tasa de acierto Masculina	Tasa de acierto Femenino	Tipo
<b>ICMIE</b>	Algoritmo de Rao izquierdo	59.4	46.0	51.9	M
<b>ICACa</b>	Árbol Clasificación	75.3	79.6	77.4	NP
<b>ICACb</b>	Árbol particionado recursivo	61.6	76.1	68.8	NP

Para el ICM (Índice Canino Mandibular) 114 mujeres de 251 y 127 hombres de 274 fueron correctamente clasificados, equivaliendo a un Índice de acierto global del 45.9% y con 43.7% en las mujeres y 48.1% en los hombres. En este mismo estudio el ICM<sub>ax</sub> (Índice canino maxilar) 109 mujeres de 215 y 131 hombres de 260 fueron correctamente clasificados, traducándose en una tasa de acierto del 50.5% global, con una tasa de acierto para las mujeres de 45,8% y de 55,3% para los hombres.

De los tres índices es indudable que a pesar de ser el ICMIE de Rao las formas más usadas, a partir de su artículo seminal, (Rao *et al.*, 1989), para el caso del estudio en Uruguay éstos muestran una capacidad predictiva muy pobre.

Por lo tanto los resultados más relevantes en este trabajo se obtuvieron usando las mismas dos variables (el DMDi y el DIC), con los árboles se incrementó notoriamente la capacidad predictiva, tal como se ve en la Tabla 8, donde es necesario aclarar que la columna Tipo se refiere a si el índice es Manual (M), y No paramétrico (NP).

La figura 5 muestra que las densidades de ambas subpoblaciones (F y M) son muy similares para el ICM<sub>axi</sub>, lo que luego se refleja en una pobre performance del ICM<sub>axi</sub>.

Los Índices que surgen de la aplicación de los métodos de CART y de particionado recursivo, son muy fáciles de visualizar e interpretar ya que son en base a umbrales de corte que surgen de las variables observadas y no consideran el tener que elaborar cálculos engorrosos posteriores como podrían ser los modelos paramétricos de Regresión Logística (RL) y de Análisis Discriminante (AD). En los 2 casos de las metodologías de CART los resultados son muy similares, mostrando ambos las mismas variables como las más importantes y con los mismos puntos de corte para la variable DMDi, que es la más importante, seguida por la DIC. Para el caso del método de CART convencional se obtienen 5 nodos terminales, donde la variable DIC con otros umbrales de corte generar nodos y por último vuelve a aparecer nuevamente DMDi.

Por el otro lado, para el caso del árbol construido por el particionado recursivo, se obtienen cuatro nodos, donde no es necesario hacer proceso de poda.

En cuanto a la performance, el ICAC<sub>a</sub>, parece tener mayor performance global pero con un sesgo muy grande hacia la tasa de acierto femenina y teniendo una mejor tasa de acierto masculina, en comparación con el ICAC<sub>b</sub>.

Por otra parte las ventajas de los Índices que podrían surgir del AD y de la RL es que permitirían ver el *peso* que tiene cada variable en la capacidad de discriminación, lo que se refleja los coeficientes de la función discriminante para el AD o en el impacto mediante los OR para el caso de la RL. La RL es un modelo paramétrico, ampliamente usado en biomedicina y para el cual se puede evaluar a través de la curva ROC, en forma más fina, su performance pudiendo encontrar el punto óptimo de corte para maximizar la

sensibilidad (Se) y especificidad (Esp), pudiendo además incorporar atributos de tipo categórico.

## 6. CONCLUSIONES

Luego de presentados tres índices que surgen de aplicar, un método convencional y otros dos basados en técnicas estadísticas presentadas mínimamente, se puede decir que para los datos del estudio, se logra una mejora muy importante en la capacidad predictiva, siendo entonces, en este caso el ICACa y ICACb, una buena alternativa y donde se recomienda dado el potencial y facilidad de interpretación, el ICACb que resulta en un árbol mediante particionado recursivo, pero sabiendo que el ICACa es superior en capacidad predictiva.

La ventaja del ICACb es que en cada rama del árbol el investigador tiene cuantificada la importancia de ese umbral de corte en términos de la prueba de hipótesis que hace en cada etapa.

Una ventaja que tiene contar solamente con dos variables explicativas en el árbol de regresión es que puede ser representada gráficamente en forma complementaria con un gráfico de dispersión, donde quedan determinadas cinco zonas para el ICACa y cuatro zonas para el ICACb. La idea es que si se grafica en forma simultánea las variables DMDi y DIC, tomando los puntos de corte de cada árbol, se establecen las diferentes zonas.

No obstante, esta mejoría encontrada, es necesario no perder de vista que se pueden mejorar la performance de los índices recomendados, si por ejemplo se usan otras las variables explicativas. En este caso hay que recordar que solo se usaron las mismas variables explicativas utilizadas por Rao en su algoritmo original, de modo de poder comparar el método convencional con las dos variantes de los modelos CART. Es esperable que usando mas variables evaluadas en el estudio, la mejoría en la predicción mediante el método CART aumente.

## Referencias

- Abernathy, J. R., Graves, R. C., Bohannon, H. M., Stamm, J. W., Greenberg, B. G., & Disney, J. A. (1987). Development and application of a prediction model for dental caries. *Community Dentistry and Oral Epidemiology*, 15, 24–28.
- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, Inc.
- Acharya, A. B. & Mainali, S. (2009). Limitations of the mandibular canine index in sex assessment. *J. Forensic Leg. Med.*, 16, 67–69.

- Acharya, A. B., Prabhu, S., & Muddapur, M. V. (2011). Odontometric sex assessment from logistic regression analysis. *International Journal of Legal Medicine*, 125(2), 199–204.
- Al-rifaiy, M. Q., Abdullah, M. A., Ashraf, I., & Khan, N. (1997). Dimorphism of mandibular and maxillary canine teeth. *The Saudi Dental Journal*, 9(1), 17–20.
- Álvarez-Vaz, R., Riaño, M., Mesa, M., Buño, G., & Nalbarte, L. (2011). Maloclusión en niños en edad escolar: Análisis de los factores de riesgo. *Colección Biblioteca Plural de la CSIC. Departamento de Publicaciones, Unidad de Comunicación de la Universidad de la República (UCUR)*.
- Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). Classification and Regression Trees. *Chapman and Hall/CRC, 1 edition*.
- Campos-Neto, M. & Paulete, J. (2014). *Atlas de medicina legal. Guia práctico para médicos e operadores do direito*, volume tomo 1. LEUD.
- Cribari-Neto, F. & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2), 1–24.
- Clark, D. (1994). An analysis of the value of forensic odontology in ten mass disasters. *International Dental Journal*, 44, 241–250.
- Costa, A. A., Pereira, M. A., Ramos, D., Meléndez, B., Silva R. F., Velos, G. S. M., & Daruge, Jr. E. (2005). Determinação do gênero por meio de medidas craniométricas e sua importancia pericial. *Revista de Medicina Legal, Direito Medico e da Saúde*, 1(3), 18-23.
- Eboh, D. & Etetafia, M. (2010). Maxillary canine teeth as supplement tool in sex determination. *Annals of Biomedical Sciences*, 9(1).
- Gargano, V., Picapedra, A., Sassi, C., Lima, L., Álvarez-Vaz, R., Francesquini Jr, L. & Daruge Jr., E. ¿Son los índices caninos mandibular y maxilar herramientas fidedignas para la determinación del sexo? *Actas odontológicas 2014; XI (1)*, 22-34.
- Harvey, J. (1975). Dental identification and Forensic Odontology, pages 140–157. John Wright & Sons, forbes, g watson aa. legal aspects of dental practice, *Bristol edition*.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674.
- Kaushal, S., Patnaik, V. V. G. & Agnihotri, G.(2003). Mandibular Canines In Sex Determination *J Anat. Soc. India*; 52(2), 119-124.
- Nahidh, M., Ali Ahmed, H., Mahmoud, A., Murad, A., & Mehdi, B. (2013). The role of maxillary canines in forensic odontology. *Journal Bagh Coll Dentistry*, 25(4), 109–113.

- Pereira, C., Bernardo, M., Pestana, D., Santos, J., & Mendonca de, M. (2010). Contribution of teeth in human forensic identification-discriminant function sexing odontometrical techniques in portuguese population. *J. Forensic and Legal Med*, 17, 105–110.
- Picapedra, A., Sassi, C., Massa, F., Francesquin Jr, L., Daruge, E., & Daruge Jr, E. (2012). Odontometric analysis of maxillas: a device for sex determination. *Inter J Dental Anthropol*, 2, 01–16.
- Prabhu, S. & Acharya, A. B. (2009). Odontometric sex assessment in indians. *Forensic Science International*, 192(1):129.e1 – 129.e5.
- Rai, B. & Anand, S. (2007). Gender determination by diagonal distances of teeth. *The Internet Journal of Biological Anthropology*, 1(110).
- Rao, N., Rao, M., Pai, L., & Kotian, S. (1989). Mandibular canine index. *Forensic Science International*
- Sassi, C., Picapedra, A., Lima, L., Francesquini Jr, L., Daruge, E. & Daruge Jr E (2012). Sex determination in Uruguayans by odontometric analysis *Braz J Oral Sci*; 11(3), 381-386.
- Srivastava, P. (2010). Correlation of odontometric measures in sex determination. *J Indian Acad Forensic Med*, 32(1), 56–61.
- Therneau, T. & Atkinson, B. (2018). *rpart: Recursive Partitioning and Regression Trees*. R package version 4, 1-13.
- Zorba, E., Moraitis, K., & Manolis, S. K. (2011). Sexual dimorphism in permanent teeth of modern greeks. *Forensic Sci. Int.*, 210(1-3), 74–81.