

Logistic Regression in Forensic Odontology

Sex determination through dental measurements

In the field of forensic odontology teeth are an excellent research material. Furthermore teeth are almost an indestructible material. The information provided by their shapes and sizes allow us to determine some characteristics of an individual such as sex or age. This research seek to evaluate the efficiency of logistic regression for sex classification. For that it was used dental pieces in a sample of 524 individuals.

Data used

The data used correspond to 524 lower plaster models(264 males and 260 females) of patients treated at an orthodontic clinic in Montevideo, Uruguay [1]. From each cast the following were obtained: the mesiodistal diameter, the gingivo-incisal height of both canines and the inter canine distance.

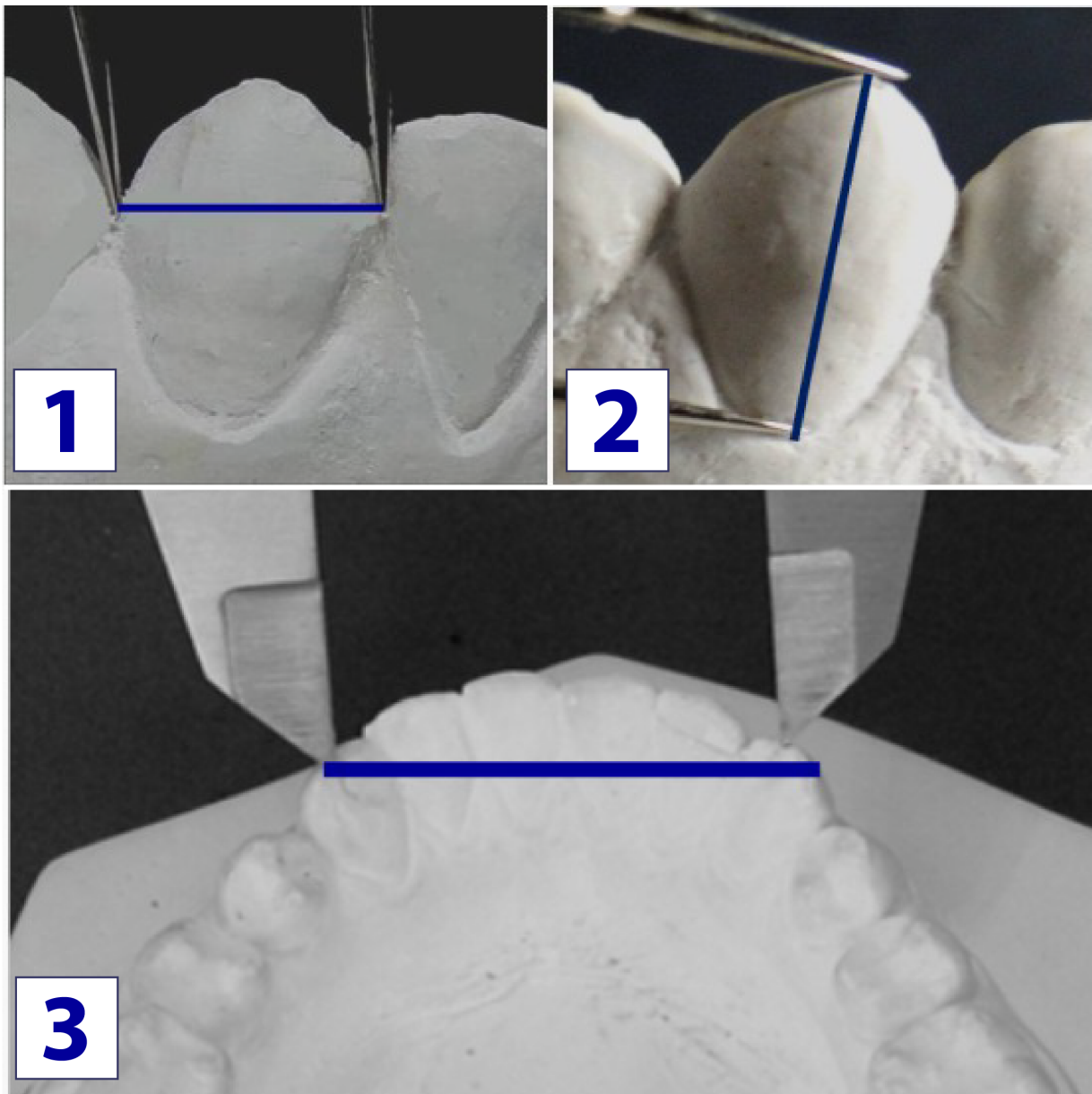


Figura 1: Considered measurements: 1 Mesodistal Diameter (DMD) 2 . Gingivo-incisal Height(AGI) 3. Inter canine Distance(DIC).

Within the set of measurements: (in milimetres), multivariate outliers were sought using Mahalanobis distance; observations whose right or left DMD exceeded 8.5 mm and whose right or left AGI exceeded 11.5 mm were removed, leaving a total of 511 observations.

Measuremet	Male (N=255)	Female (N=256)
DMD Right (DMDd)	7.110 (0.440)	6.707 (0.448)
AGI Right (AGId)	9.231 (0.957)	8.932 (0.891)
DMD Left (DMDi)	7.125 (0.465)	6.744 (0.449)
AGI left (AGIi)	9.260 (0.970)	8.977 (0.869)
DIC	26.448 (2.319)	25.491 (2.010)

Cuadro 1: Mean (SD) values by sex for each measurement (without outliers).

Methodology

First, a logistic regression model is fitted to discriminate sex by considering all measurements from both the left and the right sides. After that it was used models that only consider the measurements of each side plus the inter canine distance.

- Full Model:
 $Y : Sex; \quad X : (DMDd, DMDi, AGI, AGIi, DIC)$
- Left side model:
 $Y : Sex; \quad X : (DMDi, AGIi, DIC)$
- Right side model :
 $Y : Sex; \quad X : (DMD, AGI, DIC)$

The study focuses on the predictive power of these models and therefore evaluates their performance through several metrics: **Accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, F1 and balanced accuracy**.. All metrics are calculated by 10 fold cross validation. Analyses are performed in *R*, mainly using the *caret* library for model fitting and metric extraction

Results

The estimated coefficients and performance metrics for each model are presented.

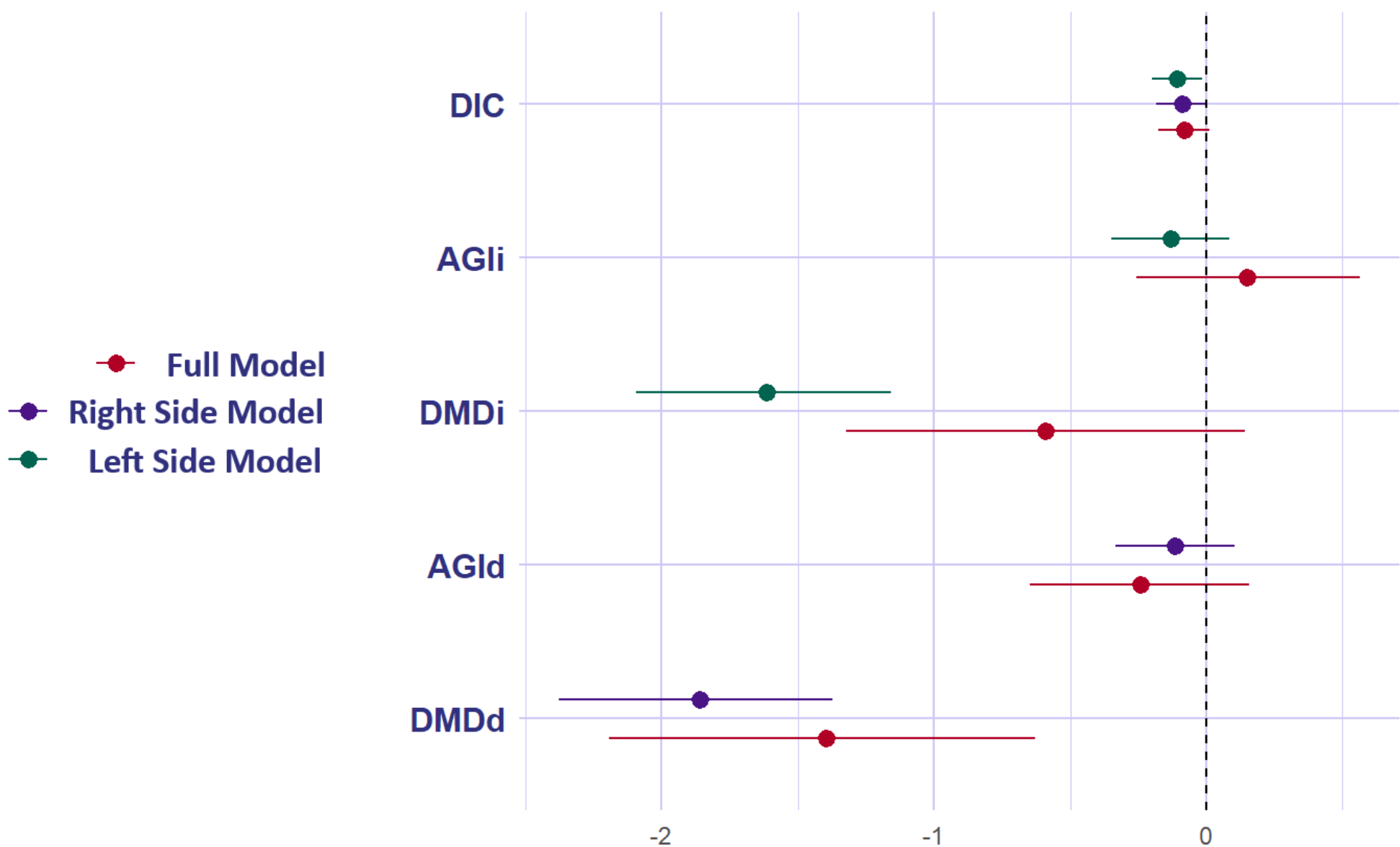


Figura 2: Estimated coefficients and 95 % confidence intervals for each model

Metrica	Full	Left side	Right side
Sensitivity	0.678	0.667	0.667
Specificity	0.684	0.668	0.711
PPV (+)	0.681	0.667	0.697
NPV. (-)	0.680	0.668	0.682
Accuracy	0.681	0.667	0.697
F1	0.680	0.667	0.681
Balanced ACC	0.681	0.667	0.689

Cuadro 2: Cross validated performance metrics.

Conclusions

Logistic regression showed good overall performance, with an accuracy of almost 70 % for the right side model. Given the sensitivity and Specificity values obtained- both around 70 % (the capacity to correctly recognize males and females, respectively) the model can distinguish the two group satisfactorily. The same conclusion is supported by the positive and negative predictive values. However, as seen in the estimated coefficients, only a few of the variables proved statistically significant, suggesting that including additional variables or constructing new indicators from the existing ones could improve the model.

References

- Alvarez-Vaz, R. Sassi, C. 2020. "Índice canino maxilar: Determinación del sexo mediante técnicas de clasificación supervisada."
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (2nd ed.)
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26.

[1] Data Source: <https://doi.org/10.60895/redata/JIU5UG>