International Journal of Health Sciences and Research
[Indexed, Peer-Reviewed Journal]
Volume 15; Issue: 6; June 2025
Website: www.ijhsr.org
Original Research Article *ISSN: 2249-9571*

Sex Differentiation Using Cephalic and Somatic Measurements: A Binary Logistic Regression Approach

Pooja Sah[1], Udai Pratap Singh[2]

[1]Research Scholar, Department of Anthropology, University of Lucknow, Lucknow, India.
[2]Professor, Department of Anthropology, University of Lucknow, Lucknow, India.

Corresponding Author: Pooja Sah

DOI: https://doi.org/10.52403/ijhsr.20250626

ABSTRACT

One of the key applications of anthropometry lies in sex estimation, particularly in forensic anthropology, where investigation of missing individuals and unidentified cadavers aids in the identification of unknown individuals. Binary logistic regression was deemed useful for sexual differentiation by developing a predictive model that estimates an individual's sex based on various variables. This model assigns predicted probabilities, and a cut-off value (typically 0.5) is used to classify individuals as male or female. The model's predictive power is often assessed using the Receiver Operating Characteristic (ROC) curve, with the area under the curve (AUC) indicating the strength of the model. Thus, it was performed to assess the ability of cephalic and somatic measurements to predict sex. This exploratory study was conducted on 131 adult population, consisting of both males and females, employing snowball sampling and purposive sampling method, using. Martin's Anthropometer, Spreading, sliding calipers and a flexible steel tape were utilised as the instrument for this study. The model was statistically significant ($p < 0.001$), and predictors including Max Head Length ($p < 0.001$), Max Head Breadth ($p < 0.001$), and Stature ($p = 0.006$) contributed significantly. Head Circumference was not statistically significant ($p = 0.456$). The odds of being female decreased with increasing values of Max Head Length, Breadth, and Stature.

Keywords: Anthropometry, Morphometry, Cephalometry, Multivariate Statistics.

INTRODUCTION

Anthropometry, a branch of morphometry, is the scientific study of the size and shape of the human body and its components. Morphometrics, more specifically, involves the quantitative analysis of biological forms, focusing on variations in size and shape among populations. Human body dimensions are influenced by ecological, geographical, racial, gender, and age factors, making anthropometry a vital tool in understanding biological diversity.

While the Anthropometric measurements are one of the most reliable tools for sex and age estimation, however, reliance on a single parameter like stature is insufficient due to substantial overlap between male and female height distributions. Thus, multiple cephalic and somatic parameters must be examined to improve accuracy in sex differentiation and to detect sexual dimorphism.

Physical anthropologists have long used anthropometric data to classify the diverse

populations of the Indian subcontinent. However, there remains a notable gap in research focused on Denotified, Nomadic, and Semi-Nomadic Tribes/ Communities that often remain underrepresented in national datasets. One such community is the Gadia Lohar of Uttar Pradesh, traditionally a semi-nomadic group known for their craftsmanship in ironwork. Originally from Rajasthan, they have migrated across multiple Indian states. Though mentioned in the 2017 Idate Commission's Report, their demographic and anthropometric data, especially for Uttar Pradesh, remains largely undocumented (1) (2)

As Franklin et al. argue, applying foreign standards often leads to substantial misclassification—particularly of females—underlining the urgency of developing localized reference data (3). Considering this, the present study undertakes an anthropometric analysis of the Gadia Lohar population of Central Uttar Pradesh, using four morphometric parameters—Maximum Head Length, Maximum Head Breadth, Head Circumference, and Stature—this research aims to assess their significance in determining sex using Binary Logistic Regression Model for accuracy. The goal is to formulate population-specific standards, which are essential for improving forensic accuracy and addressing the identity crisis faced by the Gadia Lohar amid ongoing cultural assimilation and modernization pressures.

MATERIALS & METHODS

This exploratory study was conducted on 131 adult population, consisting of both males (56) and females (75) without any physical deformities. Snowball sampling and purposive sampling methods were employed for locating the population. Instruments like the Martin Anthropometer, Spreading, sliding calipers and flexible steel tape were utilized for taking Anthropometric measurements.

A reliable and reproducible anthropometric measurement and analysis depends on the correct usage of standardized techniques. The subject should be made to stand with their natural head position in an upright posture without any strain in the eye-ear plane or F-H (Frankfort-Horizontal) plane i.e. tragion and right orbitale must lie in the same plane (4). This implies that the anthropometric measurements are more effective when visual axis is horizontal. Natural head position provides the key for meaningful analysis.

Measurements taken-

1. Maximum Head Length (Glabella-Opisthocranion): The distance from the glabella (the most prominent point on the forehead) to the most projecting point at the back of the head (the opisthocranion). This is measured with a spreading caliper.
2. Maximum Head Breadth (Euryon-Euryon): The distance between the two most lateral points on the sides of the head, typically at the parietal bones (above the ears). This measurement is also taken with a spreading caliper.
3. Head Circumference- It measures the circumference from glabella (g) to opisthocranion (op) and back to glabella (g).
4. Stature (floor-v): It measures the vertical distance from the standing floor to the vertex.

STATISTICAL ANALYSIS

The aim of this research paper is to differentiating/predicting sex, not just comparing mean differences, this is a classification problem, making logistic regression more suitable than T-test. Further to fully achieve this goal, a choice was made to utilize Binary Logistic Regression instead of Multiple Linear Regression. Along with Descriptive statistics, tests like Assessment of multicollinearity by VIF, Omnibus Test of Model Coefficients and Hosmer and Lemeshow test were also applied to check whether the variables used were appropriate or not. Significance level

was kept constant, i.e., $p < 0.05$. SPSS 25 has been utilised for analysis of the data.

RESULT

Descriptive Statistics of Variables-

The data was assessed for normality using the Shapiro-Wilk test and indicated the value of 0.067 which was within acceptable ranges, indicating normality.

Males exhibited significantly higher mean values than females across all measured variables. The mean Maximum Head Length for males was **18.98 ± 0.43 cm**, while for females it was **18.43 ± 0.45 cm** ($p = 0.000$). Similarly, the mean Maximum Head Breadth was **14.60 ± 0.48 cm** in males and **13.96 ± 0.27 cm** in females ($p = 0.000$). Head Circumference was also greater in males (**54.23 ± 0.70 cm**) compared to females (**53.70 ± 0.83 cm**), and this difference was statistically significant ($p = 0.000$). Finally, the mean Stature was significantly higher in males (**163.73 ± 5.71 cm**) than in females (**156.57 ± 6.60 cm**, $p = 0.000$). (Table 1)

Table 1 Descriptive Statistics of the Studied Population

Variables	Males (N=56)	Females (N=75)	p- value
Maximum Head Length	18.98±0.43	18.43±0.45	0.000
Maximum Head Breadth	14.60±0.48	13.96±0.27	0.000
Head Circumference	54.23±0.70	53.7±0.83	0.000
Stature	163.73±5.71	156.57±6.6	0.000

Assessment of Multicollinearity Using Variance Inflation Factors (VIFs)-

To evaluate multicollinearity among the independent variables (Max Head Length, Max Head Breadth, Head Circumference, and Stature), Variance Inflation Factors (VIFs) were computed using multiple linear regression by rotating each variable as the dependent variable. All Variance Inflation Factor (VIF) values were significantly below the standard threshold of 5, suggesting that multicollinearity does not pose a problem in this model. The predictors are sufficiently independent, justifying their inclusion in the binary logistic regression analysis. (Table 2)

Table 2 Assessment of Multicollinearity using Variance Inflation Factors (VIFs)

Dependent Variable	Predictor	Tolerance	VIF
Max Head Length	Max Head Breadth	0.942	1.061
	Head Circumference	0.879	1.137
	Stature	0.855	1.170
Max Head Breadth	Max Head Length	0.677	1.476
	Head Circumference	0.706	1.416
	Stature	0.821	1.213
Head Circumference	Max Head Length	0.844	1.185
	Max Head Breadth	0.942	1.061
	Stature	0.823	1.213
Stature	Max Head Length	0.721	1.388
	Max Head Breadth	0.967	1.034
	Head Circumference	0.723	1.383

The Omnibus Test of Model Coefficients-

It was used to determine whether the set of morphometric predictors significantly improves the model's ability to differentiate sex compared to a model with no predictors (i.e., only the intercept). The test yielded a Chi-square value of 121.889 with 4 degrees of freedom, and the result was statistically significant ($p < 0.001$).

This indicates that the inclusion of the predictor variables—Maximum Head Length, Maximum Head Breadth, Head

Circumference, and Stature—significantly contributes to predicting the sex of individuals.

Therefore, the binary logistic regression model provides a significantly better fit to the data than a null model without predictors.

Table 3 Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	121.889	4	0.000
	Block	121.889	4	0.000
	Model	121.889	4	0.000

The Hosmer and Lemeshow Test-

It was used to assess the goodness of fit for the binary logistic regression model. By using predictors such as Maximum Head Length, Maximum Head Breadth, Head Circumference, and Stature, it was determined that with p = 0.662 (which is significantly greater than 0.05), our data fits within the statistically appropriate model for classification. (Table 4)

The model summary indicates a -2 Log Likelihood value of 56.950, which reflects how well the logistic regression model fits the data. Lower values suggest a better fit.

The Cox & Snell R^2 = 0.606 and the Nagelkerke R^2 = 0.813 provide pseudo-R^2 values, which estimate the proportion of variation in the dependent variable (sex) explained by the model.

In particular, the Nagelkerke R^2 value of 0.813 suggests that approximately 81.3% of the variability in sex classification can be accounted for by the morphometric predictors used in the model—Maximum Head Length, Maximum Head Breadth, Head Circumference, and Stature. This indicates a very strong model fit and supports the use of binary logistic regression over simpler tests like the t-test when considering multiple predictors simultaneously. (Table 5)

Table 4 Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.870	8	0.662

Table 5 Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	56.950[a]	0.606	0.813
a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.			

The following contingency table demonstrates that observed frequencies are generally close to expected frequencies across deciles of predicted probabilities, indicating [a good/poor] agreement between predicted and observed classifications. (Table 6)

Table 6 Contingency Table for Hosmer and Lemeshow Test

		Gender = 1		Gender = 2		
		Observed	Expected	Observed	Expected	Total
Step 1	1	13	12.984	0	0.016	13
	2	13	12.868	0	0.132	13
	3	12	12.292	1	0.708	13
	4	10	10.344	3	2.656	13
	5	7	4.214	6	8.786	13
	6	0	2.044	13	10.956	13
	7	1	0.888	12	12.112	13
	8	0	0.275	13	12.725	13
	9	0	0.078	13	12.922	13
	10	0	0.012	14	13.988	14

Sex Determination using Binary Logistic Model-

Table 7 demonstrates the result of Binary logistic regression, which was conducted to evaluate the ability of cephalic and somatic measurements to predict sex. The model was statistically significant ($p < 0.001$), and predictors including Max Head Length ($p < 0.001$), Max Head Breadth ($p < 0.001$), and Stature ($p = 0.006$) contributed significantly. Head Circumference was not statistically significant ($p = 0.456$). The odds of being female decreased with increasing values of Max Head Length, Breadth, and Stature. It should be mentioned here that the model included a statistically significant intercept ($B = 204.292$, $SE = 48.767$, $p < 0.001$), which represents the baseline log-odds of sex classification when all predictor variables are zero (Table 7). Although biologically implausible for human morphometric values to be zero, the intercept is mathematically required for the logistic regression model to compute predicted probabilities accurately.

Table 7 Demonstrating the Variables in the Binary Logistic Regression Model

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Maximum Head Length	-3.766	1.020	13.640	1	0.000	0.023
	Maximum Head Breadth	-5.665	1.189	22.697	1	0.000	0.003
	Head Circumference	-0.448	0.602	0.556	1	0.456	0.639
	Stature	-0.179	0.066	7.442	1	0.006	0.836
	Constant	204.292	48.767	17.549	1	0.000	5.28445E+88
a. Variable(s) entered on step 1: Maximum Head Length, Maximum Head Breadth, Head Circumference, Stature.							

The Classification Table from the Binary Logistic Regression analysis shows that the model has classified:

- **48 out of 56 males** with 85.71% accuracy and
- **69 out of 75 females** with 92.0%.

And the overall **prediction accuracy is 89.3%**, which is excellent for a classification model in anthropological or forensic research. The cut-off probability value used was **0.5**, meaning: Individuals with predicted probability ≥ 0.5 were classified as **Female (2)** and those < 0.5 were classified as **Male (1).** (Table 8)

Table 8 Classification Table

Observed			Predicted		
			Gender		
			1	2	Percentage Correct
Step 1	Gender	1	48	8	85.71
		2	6	69	92.0
	Overall Percentage				89.3
a. The cut value is .500					

The classification plot presents the distribution of observed and predicted group memberships (sex) based on the logistic regression model. The x-axis represents the **predicted probability of being female (Group 2)**, while the symbols '1' and '2' correspond to the actual group membership—1 for male and 2 for female. The majority of male cases (Group 1) are clustered at lower predicted probabilities (left side of the plot), whereas female cases (Group 2) are concentrated toward higher predicted probabilities (right side of the plot). (Graph 1)

This clear separation between groups suggests that the model is effectively distinguishing between male and female individuals based on the predictor variables. The cut-off value used was **0.50**, meaning predicted probabilities above 0.50 were

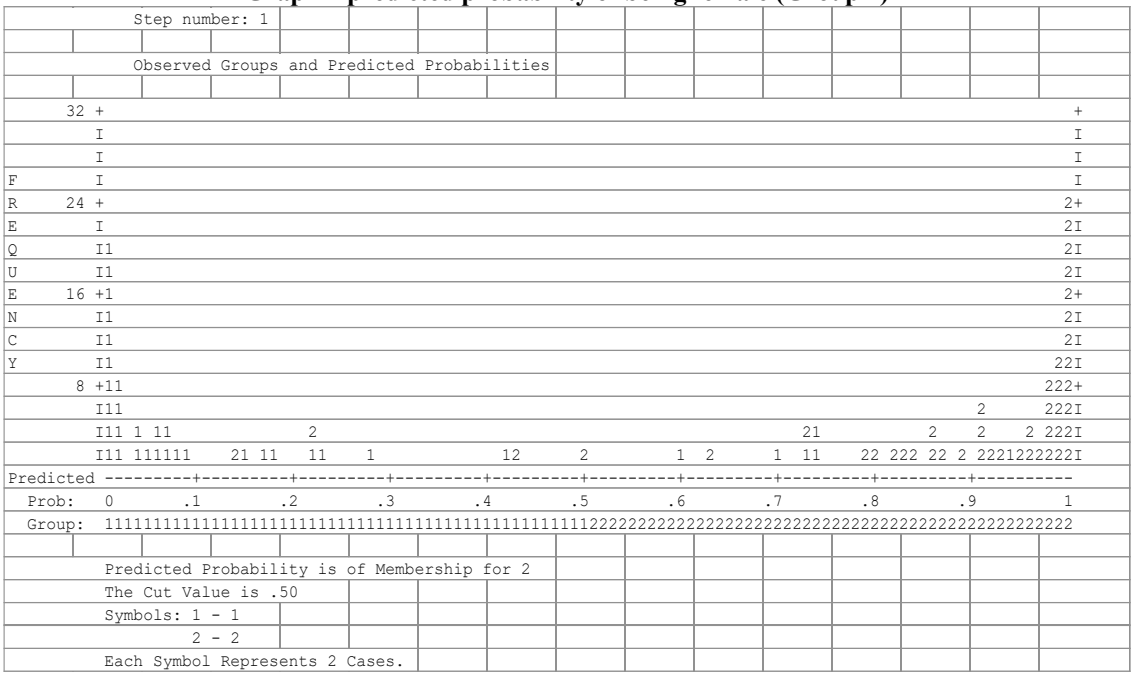
classified as female, and those below as male.

The horizontal display at the bottom shows a clear separation between groups:

- Group 1 (Females) mostly cluster between 0.0 to 0.4, meaning the model predicts they are unlikely to be male.
- Group 2 (Males) mostly cluster between 0.6 to 1.0, meaning the model predicts them correctly as male.
- Very few individuals fall near 0.5, where classification is ambiguous.

This aligns with the high classification accuracy (89.3%) and non-significant Hosmer-Lemeshow test (p = 0.662) — both indicate excellent model fit. And suggests strong discriminative power of the logistic model. Misclassifications were minimal and typically occurred near the decision boundary of 0.5.

Graph 1 predicted probability of being female (Group 2)



Receiver Operating Characteristic (ROC) Curve Analysis-

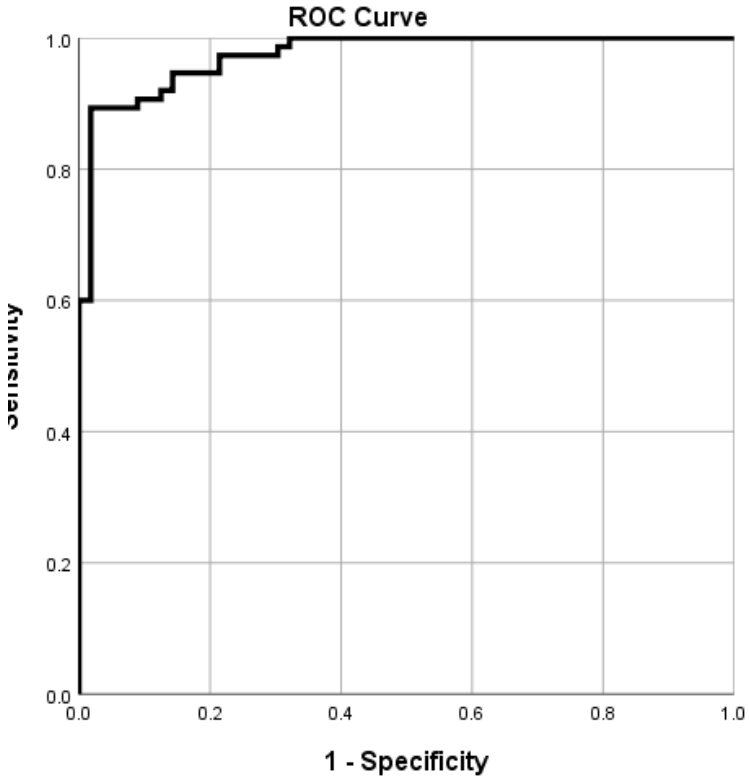
The ROC (Receiver Operating Characteristic) curve is depicted in (Graph 2) to evaluate the discriminatory performance of the binary logistic regression model. The curve illustrates the trade-off between sensitivity (true positive rate) and 1 - specificity (false positive rate) across a range of classification thresholds.

In this analysis, the area under the curve (AUC) was approximately very high, suggesting excellent predictive ability of the model in distinguishing between the two groups (e.g., male and female, depending on coding). The curve of the plot reflects strong specificity and sensitivity.

This strong classification performance supports the validity of the selected anthropometric predictors in differentiating sex within the studied sample. The ROC curve thus confirms that the logistic regression model is robust and reliable for binary classification based on the included variables.

The area under the ROC curve (AUC) was 0.96, indicating a high level of model accuracy.

Graph 2 Receiver Operating Characteristic (ROC) Curve Analysis



DISCUSSION

The present study aimed to assess sexual dimorphism among the adult population of Gadia Lohar of Central Uttar Pradesh, using selected cephalic and somatic anthropometric variables—Max Head Length, Max Head Breadth, Head Circumference, and Stature—through Binary Logistic Regression.

The descriptive statistics revealed that there was a strong sexual dimorphism in the chosen morphometric traits, with males consistently showing higher mean values than females, like previous other studies. (5) The statistically significant *p*-values (< 0.05) obtained through independent samples *t*-tests also provide support to the conclusion that gender differences are highly significant for all variables (6) (7–9)

Binary Logistic Regression model was utilised to verify and check the best accuracy level for sex determination., which reveals statistically significant differences between male and female individuals, with Max Head Length, Max Head Breadth, and Stature emerging as robust predictors of sex ($p < 0.001$ for both head dimensions and $p = 0.006$ for stature). Head Circumference, however, did not contribute significantly ($p = 0.456$), possibly due to its lower variability or shared overlap in values between sexes within the studied population.

The inverse relationship observed between increasing values of Max Head Length, Breadth, and Stature and the likelihood of being female aligns with established anthropological literature. Males generally exhibit greater cephalic dimensions and overall stature than females due to genetic, hormonal, and functional differences rooted in sexual selection and developmental biology (10–14). These differences, although population-specific, are consistent with known sexual dimorphism in humans.

The high classification accuracy 89.3%, with 85.71% of males and 92.0% of females correctly classified, underscores the discriminative strength of the model. Such accuracy levels are considered excellent within anthropological and forensic contexts, especially when using non-

invasive, external anthropometric measures. Importantly, the non-significant Hosmer-Lemeshow test ($p = 0.662$) and high AUC value (0.96) further validate the model's goodness-of-fit and predictive capacity, affirming its reliability in sex estimation for this population. Also, our model achieved an AUC of 0.813, indicating strong discrimination between male and female subjects based on the selected morphometric parameters.

The clear separation of predicted probabilities between sexes in the classification plot highlights the practical utility of this model in field applications, such as forensic identification or bioarchaeological analysis. The small number of misclassifications, concentrated near the 0.5 threshold, suggests that while the model performs well overall, borderline cases may still require additional metrics or contextual information for more accurate classification.

The non-significance of Head Circumference invites further reflection. While head circumference is frequently employed in paediatric growth studies and developmental assessments, it may be less effective in distinguishing adult sexual dimorphism. This variable's limited predictive power may reflect a more complex interaction of cephalic vault morphology, age-related changes, and potential cultural influences on headgear use or cephalic deformation practices—factors worth exploring in future anthropological investigations.

From a broader perspective, the findings affirm that standard anthropometric variables—when applied judiciously—remain powerful tools in biological anthropology. They provide quantifiable, replicable means of assessing population-level variation and individual identity. In the context of the studied group, these morphometric patterns may also reflect the effects of localized gene flow, occupational behaviour, or adaptive responses, which should be explored in conjunction with ethnographic or genetic data for a more holistic understanding.

Overall, the binary logistic regression model developed in this study demonstrates a high degree of utility and robustness in distinguishing sex based on cephalic and somatic features. The implications extend beyond academic inquiry to include forensic anthropology, identity reconstruction, and even population health monitoring, especially in resource-limited settings where advanced imaging tools are unavailable.

The current study has some strengths: sex estimation from the cephalic measurements and stature (height vertex) are essential for medico-legal investigations globally, thinking from the perspective of forensic anthropologists. (13) (15) However, such models are population-specific due to genetic and environmental variabilities.

No existing model based on cephalic and post-cephalic measurements for sex estimation has been developed for the Gadia Lohar community of Central Uttar Pradesh. This research is therefore a pioneering effort that contributes valuable new data to the limited body of knowledge on this population. However, the authors acknowledge that the findings of this study cannot be generalized for the entire Gadia Lohar population due to inter-population variabilities. Further studies should focus on specific cultural groups in the Gadia Lohar population.

CONCLUSION

The Binary Logistic Regression model correctly classified 85.7% of male cases and 92.0% of female cases, yielding an overall accuracy of 89.3%. This high predictive accuracy suggests that the selected cephalic and somatic variables—Maximum Head Length, Maximum Head Breadth, Head Circumference, and Stature—are effective discriminators for sex determination in the studied Gadia Lohar population. (Table 7) The study highlighted that, if the assumptions of the statistical methods are met, both methods are equally capable of discriminating between the population based

on sex. The study suggests that the binary logistic regression model should be employed in forensic anthropological studies.

Declaration by Authors
Ethical Approval: Approved
Acknowledgement: None
Source of Funding: None
Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. Idate BR. National Commission for Denotified Nomadic and Semi-Nomadic Tribes. Ministry of Social Justice & Empowerment, Government of lndia; 2017 Dec p. 276.
2. Sah P, Singh UP. Exploring Blood Group Variability in the Gadia Lohar Community of Lucknow, Uttar Pradesh. Antrocom Journal of Anthropology. 2024;20(02):51–9.
3. Franklin D, Cardini A, Flavel A, Kuliukas A. Estimation of sex from cranial measurements in a Western Australian population. Forensic Science International. 2013 Jun 10;229.
4. Rastogi S, Shukla BRK. Laboratory Manual of Physical Anthropology: Anthropometry and Osteology. 1st ed. Lucknow: Bharat Book Centre; 2017.
5. Shah T, Patel M, Nath S, Menon S. Determination of sex using cephalo-facial dimensions by discriminant function and logistic regression equations. Egyptian Journal of Forensic Sciences. 2015 Apr 29;147.
6. Kumar M, Lone MohM, Patnaik VVG. Determination of sex by discriminant function analysis: A cephalometric study. International Journal of Pure & Applied Bioscience. 2013;01(04):18–21.
7. Ilayperuma I. Evaluation of Cephalic Indices: A Clue for Racial and Sex Diversity. International Journal of Morphology. 2011 Mar;29(1):112–7.
8. Sol M del. Cephalic Index in a Group of Mapuche Individuals from the IX Region of Chile. International Journal of Morphology. 2005 Jan 1;23(03):241–6.
9. Mahajan A, Khurana BS, Seema, Batra APS. The study of cephalic index in Punjabi students. Journal of Punjab Academy of Forensic Medicine & Toxicology. 2009; 09(02):66–70.
10. Eboh DEO, Ohaju-Obodo JO. Height estimation from head dimensions in South-South Nigeria | Egyptian Journal of Forensic Sciences | Full Text. Egyptian Journal of Forensic Sciences. 2019;09.
11. Maina M, Mahdi O. Sexual dimorphism in cranial dimensions among three ethnic groups of North-Eastern Nigeria. American Journal of Scientific and Industrial Research. 2011;2(6):871–6.
12. Orish CN. Cephalic Index in Sexual Dimorphism and Racial Diversity: A Mini Review. MOJ Anatomy & Physiology. 2018 Jan 19;5(1).
13. Oladipo GS, Okoh PD, Akande PA, Oyakhire MO. Anthropometric study of some craniofacial parameters: head circumference, nasal height, nasal width and nasal index of adult Omoku indigenes of Nigeria. American Journal of Scientific and Industrial Research. 2011;02(01):54–7.
14. Oladipo GS, Olotu JE. Anthropometric comparison of cephalic indices between the Ijaw and Igbo tribes. Global Journal of Pure and Applied Sciences. 2006 Jun 15;12(1):137–8.
15. Oladipo GS, Okoh PD, Isong EE. Anthropometric Studies of Cephalic Length, Cephalic Breadth and Cephalic Indices of the Ibibios of Nigeria. Asian Journal of Medical Sciences. 2010;02(03):104–6.

How to cite this article: Pooja Sah, Udai Pratap Singh. Sex differentiation using cephalic and somatic measurements: a binary logistic regression approach. *Int J Health Sci Res.* 2025; 15(6):199-207. DOI: *10.52403/ijhsr.20250626*

******