Morphometric Analysis of Cranial Shape

Thomas Campbell Arnold & Dr. Dennis E. Slice
Department of Scientific Computing, Florida State University, Tallahassee, Florida

Introduction

Advances in technology have brought about a revolution in data acquisition techniques. Previously, most data collection was done via direct acquisition. Measurements would be recorded directly from each specimen, and it could take months or even years to build a substantial sample size. If an important measurement was later discovered, an entirely new data set would have to be collected. Now, laser-based surface scanners can digitize individuals in less than a minute. Thousands of data points construct the digital image, which can be stored in a library among other individuals. If an important measurement is later determined, it can easily be obtained from the digitized images. This indirect method of data acquisition has many advantages. However there are still issues preventing indirect acquisition from being utilized universally.

One such area of disruption is making measurements that are obscured by hair or other surface features. For example, surface scanners cannot see beyond hair in order to give accurate measurements of the back of a person's head. In this project, this issue is addressed by attempting to predict certain key cranial (head) measurements based on facial feature that can easily be distinguished in the scans.

Materials and Methods

Data cleaning – supplementing the average

The data set used in the study consisted of ninety-six bodily measurements for over four thousand individuals provided by the U.S. Army as part of a contract to Dennis E. Slice. Three additional measurements were calculated based on the original measurements. They were the Body-Mass Index, Cranial Index, and Facial Index.

$$BMI=\frac{(Weight)}{(Height)^2} \qquad Facial\ Index=\frac{(Cheek\ Breadth)}{(Face\ Height)}x100 \qquad Cranial\ Index=\frac{(Cranial\ Breadth)}{(Cranial\ Length)}x100$$

While the majority of individuals had all ninety-six measurements, some individuals were missing measurements. If individuals were missing either the cranial breadth or length measurement, which are needed to calculate the cranial index, they were removed from the data set. If an individual had many missing measurements then they were also removed. If an individual was only missing a few measurements then the sample average was substituted for the missing value. This method does not accurately represent the missing measurements for an individual, but it allows the individual to be included in the study by providing a reasonable estimate of the missing value. After the cleaning and repair three thousand eight hundred and thirty three individuals were included in the data set. The data was stored in Excel (Microsoft) or OpenOffice (www.openoffice.org) Spreadsheets. The data cleaning was done in R (www.r-project.org).

Correlation – R squared values

To determine which measurements have high predictive value with respect to the cranial index, the R squared value was utilized. The R squared value is a statistical measurement of a model's prediction accuracy in relation to the variability of the data. Initially, the R-squared value was calculated for each measurement with each of the other ninety-eight measurements. The measurements with high R-squared value with the cranial index, breadth, and length were noted and linear models of these measurements were graphed. The average error for each of these models was calculated in order to see which models were the most accurate predictors.

Modeling – linear regressions

A linear regression between each measurement and cranial index was carried out. Measurements such as cranial breadth, length, and circumference were excluded from the regression process because they cannot be determined from the surface scans. The leave-one-out (LOO) method of predictive accuracy was used to assess the various models. LOO consists of constructing a model using all but one data point. The models is then used to predict this data point, which has had no influence on the model. Using the LOO method, the top five most accurate models for predicting the cranial index of each individual were recorded.

Results

The facial index is defined as the bizygomatic (cheek) breadth divided by the menton to sellion (face height) length multiplied by one hundred. The cranial index is the cranial breadth divided by cranial length multiplied by one hundred. The average of the cranial index was 77.61907 and the average facial index was 116.8951. The millimeter units of each measurement cancel in their ratio making both the cranial and facial index unitless.

The facial index had the highest correlation with the cranial index out of the ninety-four variables tested from the data set. The linear regression of the facial index and the cranial index had an R squared value of 0.1338 (explaining about 13% of the variance of cranial index) and was determined to have a statistically significant correlation. The trend-line in the plot indicates that generally the facial index and cranial index increase and decrease together.

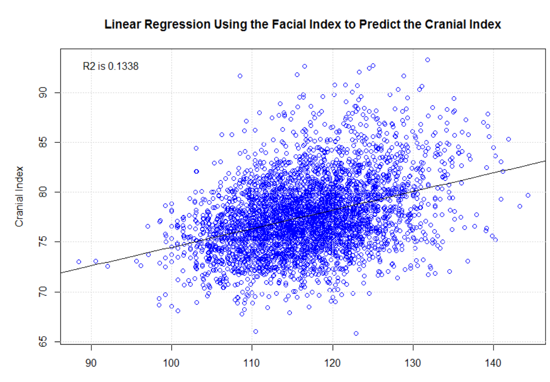

Figure 1 – Scatter plot of the Facial Index (x-axis) and the Cranial Index (y-axis) showing the linear regression line and the R-squared value.

The linear model in Figure 1 includes all points in the data set. However, the (LOO) method was used to test the effectiveness of the model without using data points utilized in the model's construction. The root mean squared error for the LOO models was 3.47 and the average error without the LOO method was 3.46; the close numbers are not surprising because the sample size is so large. The influence one data point has on the model is relatively small when there are nearly four thousand other data points. The variance of the model is 12.0787, which is almost exactly the squared LOO root mean squared error. The variance in relation to the average value is called the coefficient of variance; a lower value corresponds with a more effective model. In this case the coefficient of variance is 0.048127.

The percent confidence was calculated by placing the LOO prediction errors in order and checking the value at every five percent interval. In order to put the error in perspective, the error divided by the average value is also provided. This shows the percent error from the average.

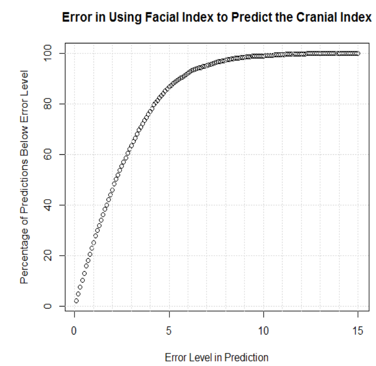

Figure 2 – Plot of what percentage individuals fall within an error level as your error increases.

Percent Confidence	Error Range	Error/Average Value
100	15.6183670	0.201218167
95	6.9586152	0.089650845
90	5.5859652	0.071966402
85	4.7997265	0.061836950
80	4.2505628	0.054761836
75	3.8369945	0.049433658
70	3.4432365	0.044360703
65	3.0961316	0.039888801
60	2.7749236	0.035750539
55	2.4735694	0.031868062
50	2.1890159	0.028202037
45	1.9445595	0.025052600
40	1.7001220	0.021903405
35	1.4419113	0.018576766
30	1.2023181	0.015489983
25	0.9938883	0.012804692
20	0.7874042	0.010144468
15	0.5713480	0.007360923
10	0.3869386	0.004985096
5	0.2106507	0.002713904

Figure 3 – Raw data values for points plotted in figure 2.

Conclusion

The facial index can readily be calculated from a surface scan in most subjects. Using this measurement as a linear model for predicting the cranial index could be very useful, since the cranial index is often obscured by hair. The linear model based on the facial index was able to make predictions with a max error range of under seven with ninety-five percent confidence. This number reflects the limitations of a linear model, which is based on just a single variable. In order to increase the accuracy other measurements that have correlation with the cranial index could be utilized in a multiple linear regression. Hopefully the introduction of additional predictors will allow the model to be even more accurate.

Based on the top five ranking of the linear models of all measurements, ear protrusion would seem a likely candidate for a second predictor. Ear protrusion had the second highest predictive power; however, there are key issues with this to be considered. While ear protrusion can be measured from most surface scans, it is often obscured by hair much like the other cranial measurements. Importantly, too preliminary analysis shows that the variation ear protrusion contributes is largely absorbed by the facial index.

Acknowledgments

The authors express gratitude to Kathryn O'Donnell, K. James Soda, Benjamin Pomidor, Detelina Stoyanova, and Cameron Berkley; the members of the FSU Mophometrics Lab, as well as the Department of Scientific Computing. Additional thanks go to Carleen Mae Sabusap, Joe O'Shea, and the member of the Office of Undergraduate Research.

This work was supported by an undergraduate research fellowship to Thomas Campbell Arnold by the Office of Undergraduate Research.