teeth!

Introduction

This paper aimed to compare the teeth of two groups of animals named scriptus and pricei. The primary objective was to conduct hypothesis testing to determine whether the means of their tooth shapes are statistically different.

Methods

Data Collection and Importation:

The data was collected in a jpg format. Black and white teeth images were imported to R using the function import_BW. This function takes a file path as an argument and uses import_jpg to load the image. Then, the lapply function was employed to apply the import function to a list of file paths (file_list_BW_extant), resulted in a list of BW images named teeth_BW_train. The names of these images were extracted from the file paths, adjusted to remove file extensions, and assigned to the images in the list. The file names were extracted from the paths in file_list_BW_extant using strsplit and were assigned to the imported images. The names of the images were modified by removing the last four characters, assuming that these are file extensions (e.g., ".jpg").

Data Preprocessing:

We had to prepare the data to use it in MATLAB. Function make_same_num_points was used to resample a curve to have the same number of points (N = 500), and then was applied to each element in specific parts of the data list. Then we expanded it to all the other images by a for loop. The prepared data was then stored into a new list named data_for_matlab, with separate matrices for each "scriptus" and "pricei" group and was saved as a csv format file.

In MATLAB, cell arrays for tooth types and species were defined. Data was read from the specified CSV file. Variables for storing processed data, including an array named teeth_data for the coordinates of each tooth, were defined. For loops were used through the data, to extract coordinates, and resamples the data to have 100 points for each tooth. The shape means were calculated, and the results were saved in the different file formats for later use.

Statistical Analysis:

For the analysis part, we again used R. As always, data was loaded.

Hotelling's T-Squared Test:

The means of tooth shapes from scriptus and pricei were compared using Hotelling's T-squared test. This multivariate statistical test was chosen for its effectiveness in detecting differences in shape means between

the two groups. Therefore, a grouping variable named g was Created where 1 represented "scriptus" and 2 represents "pricei". The length of each group was determined by the length of the corresponding lists in the "scriptus" and "pricei" data. We conducted Hotelling's T-squared test using the hotelling.test function in the Hotelling library.

Also, permutation testing was employed to assess the significance of the observed Hotelling's T-squared statistic. This is a non-parametric approach which involves randomly permuting the group labels and recalculating the test statistic multiple times. The p-value was then calculated based on how often the permuted statistic was greater than or equal to the observed statistic.

Distance Based Test:

For the better accuracy, the distance-based permutation test was also conducted. Pairwise distances between teeth were computed in MATLABand was used to compute the test statistic and a permutation testing was performed to obtain a distribution of the test statistic. The p-value was calculated based on how often the permuted statistic is greater than or equal to the observed statistic.

Principal Component Analysis (PCA) of Individual Means:

Principal Component Analysis (PCA) was applied to the individual means of each tooth type. For this section, a subset of 10 points was chosen. TBD

Results

Conclusions

Acknowledgements

Supplementary Material

References