Wolf Morphometric Data Analysis

Sujay Ravi
sravi2@binghamton.edu

***Abstract* - We are analyzing morphometric data of wolves which contains the nine morphometric parameters like the palatal length, post palatal length, crown length, etc. of two different breeds of endangered wolf species and with two gender classifications. We are using MANOVA to find how differentiable the groups are, followed by the discriminant analysis to significantly discriminate the variables which affect the species classification. We use this data to help animal researchers classify these two wolf breeds which may help them in undergoing further research on these endangered wolf species to save them from extinction.**

Index Terms **-** Discriminant analysis, MANOVA, Morphometric data of Wolves, Rocky mountain and arctic.

INTRODUCTION

Wolves are mammals which were once hunted in large scale and are still listed as endangered species in several countries. The wolf population in the south has been declining in the southern areas throughout the world and the enlistment of these species as endangered in laws throughout the world has been helpful in increasing the wolf population in the recent years [1]. Though there is increase in the wolf population there is still as significant amount of hunting going on in many parts of the world. There are also other causes of death of the wolves. To classify the wolves based on the morphometric data can give the wildlife researches, scientists and other wildlife conservation activists an idea on which wolf it was and also provide data and sometimes evidence on the death and hunting of the wolves. The data is of the skull morphometrics of the two species of the wolf which was collected to perform multivariate analysis to find the difference in the species and gender [2].

DATA

Data was collected from the Skull morphometric data on Rocky Mountain and Arctic wolves (Canis Lupus L.) taken from Morrison (1990), 3rd edition, p. 288-289, which was taken from Jolicoeur (1959, 1975) [2]. The data contains 9 independent variables. The data was about the skull morphology of wolves. The skull data had the following independent variables. The data measurements are in millimeters in this dataset.

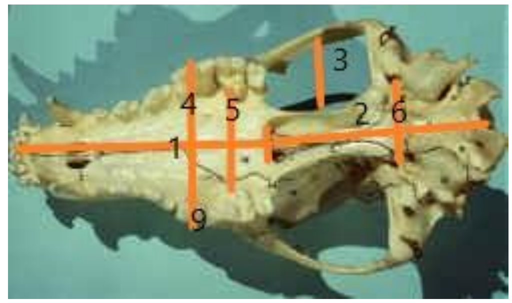

FIGURE I
WOLF SKULL MORPHOMETRICS - PALATAL VIEW [3]

1. **Palatal length:** It is the length palate of upper side of the mouth above tongue of the wolf skull.
2. **Post palatal length:** It is the length of the skull from the palate to the back of the skull.
3. **Zygomatic width:** It is the width of the zygomatic arch of the wolf skull which is usually around 12- 14 cm.
4. **Palatal width outside first upper molars:** It is the palatal width between the outer ends of the first molar teeth. Usually there are 3 molar teeth found in the wolf.
5. **Palatal width inside second upper molars:** It is the inner palatal width between the middle upper molars on the upper side of the skull.
6. **Width between the postglenoid foramen:** Foramen is an important part of the skull which can accomodate the nerves and other tissue. There are few foramina in the skull and the postglenoid foramen is the one near the brain case, behind the zygomatic bone on the either side.
7. **Interorbital width:** Orbits are the eye sockets. It is the distance between the two orbits of the skull.
8. **Brain case width:** It is the smallest width of the brain case of the wolf skull.
9. **Crown length:** It is the crown length of the first upper molar. The length of the teeth beyond the gums and it is the exposed part of the teeth.

The continuous data is in two columns one being the location of the wolf (breed type) it is of two categories, the Arctic wolf and the Rocky mountain wolf. The other dependent variable is the sex of the wolf, the male and the female respectively giving rise to four different categories of classification.

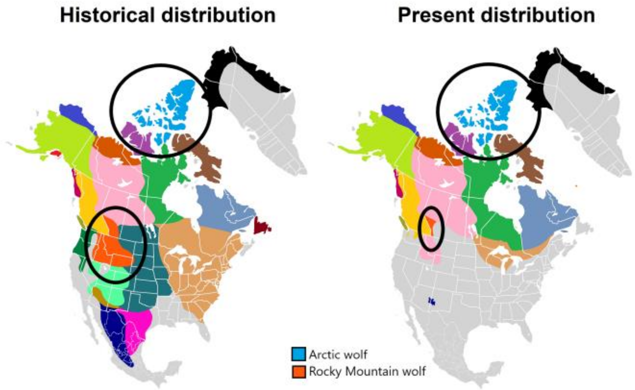

FIGURE II
WOLF DISTRIBUTION - HISTORIC VS PRESENT [4]

SOFTWARE USED

SAS and SPSS were used in a combined fashion to perform the analysis of the data.

- SPSS was used to perform MANOVA and discriminant analysis.
- SAS was used along with the SPSS to perform cross-verification of discriminant analysis to the SPSS output and also to obtain few other analysis results in a clear fashion.

MANOVA

We now try to check the relationship between the two subspecies of wolves using MANOVA. Before proceeding into MANOVA analysis we check for assumptions like multivariate normality and homogeneity.

Box's M		13.471
F	Approx.	1.050
	df1	10
	df2	1273.607
	Sig.	.398

FIGURE III
BOX'S M - TEST RESULTS

From the above Box's M test, we see that the p value is greater than 0.05. This means that it is not significant. Hence we can conclude that we fail to reject the null hypothesis, which indicates that the data have equal covariance matrices.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
x1	.204	25	.009	.928	25	.076
x3	.149	25	.156	.965	25	.533
x2	.117	25	.200*	.978	25	.837
x4	.134	25	.200*	.967	25	.561
x5	.067	25	.200*	.977	25	.826
x6	.091	25	.200*	.956	25	.339
x7	.104	25	.200*	.974	25	.746
x8	.116	25	.200*	.961	25	.442
x9	.127	25	.200*	.975	25	.776

*. This is a lower bound of the true significance.

FIGURE IV
TEST OF NORMALITY

From above shapiro-wilks test (as data points are less than 50) we can conclude that the data is multivariate normal as the significance level > 0.05.
So, as the two test meet the assumptions of MANOVA, we can perform MANOVA. We have taken the dependent variables as the nine variables i.e. the morphometrics and the independent variables being sex and location. The results of MANOVA is based on one two-way interaction and two main effects location and sex.

```
Multivariate Tests of Significance (S = 1, M = 3 1/2, N = 5 1/2)

Test Name        Value     Exact F    Hypoth. DF    Error DF    Sig. of F

Pillais          .64944    2.67596      9.00        13.00        .052
Hotellings      1.85259    2.67596      9.00        13.00        .052
Wilks            .35056    2.67596      9.00        13.00        .052
Roys             .64944
Note.. F statistics are exact.
```

FIGURE V
INTERACTION EFFECT - LOCATION * SEX

From the results of the multivariate test of significance of the two way interaction between the location and sex shows that the p value is 0.052 for all the three types of test using Pillai's trace, Hotelling trace and Wilk's lambda. Since the p value is greater than 0.05 the univariate test is not taken into account, because of the interaction between the location and sex being not statistically significant. But from practical standpoint, we can still consider the interaction effect as the p value of the test is very close to 0.05, one can say that there could be a small amount of interaction between location and sex.

```
Multivariate Tests of Significance (S = 1, M = 3 1/2, N = 5 1/2)

Test Name        Value     Exact F    Hypoth. DF    Error DF    Sig. of F

Pillais          .86537    9.28431      9.00        13.00        .000
Hotellings      6.42760    9.28431      9.00        13.00        .000
Wilks            .13463    9.28431      9.00        13.00        .000
Roys             .86537
Note.. F statistics are exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Multivariate Effect Size and Observed Power at .0500 Level

TEST NAME     Effect Size     Noncent.     Power

 (All)          .86537        83.55883      1.00
```

FIGURE VI
EFFECT OF SEX - MULTIVARIATE TEST OF SIGNIFICANCE AND POWER

Based on the multivariate statistics of the first main effect sex, we say that it is statistically significant as the p value of all the three test results is smaller than 0.05. It also shows that the power of the multivariate test is high, indicating that when the null hypothesis is false the probability of rejection is high. Since the multivariate effect is significant here, the next step is to identify the key variables which are responsible for the difference using the univariate significance test.

EFFECT .. SEX (Cont.)
Univariate F-tests with (1,21) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F	ETA Square
palatal_	260.54493	100.60000	260.54493	4.79048	54.38811	.000	.72144
postpala	133.93623	165.93333	133.93623	7.90159	16.95055	.000	.44665
zygomati	625.57826	557.90000	625.57826	26.56667	23.54749	.000	.52859
palata_1	107.14436	89.31767	107.14436	4.25322	25.19134	.000	.54537
palata_2	17.34545	55.90933	17.34545	2.66235	6.51509	.019	.23678
width_be	45.19936	108.62933	45.19936	5.17283	8.73785	.008	.29383
interorb	33.57223	159.30067	33.57223	7.58575	4.42570	.048	.17406
least_wi	4.17926	138.48900	4.17926	6.59471	.63373	.435	.02929
crown_le	4.61009	5.25267	4.61009	.25013	18.43099	.000	.46742

Variable	Noncent.	Power
palatal_	54.38811	1.00000
postpala	16.95055	.97488
zygomati	23.54749	.99617
palata_1	25.19134	.99767
palata_2	6.51509	.68033
width_be	8.73785	.80320
interorb	4.42570	.51739
least_wi	.63373	.14794
crown_le	18.43099	.98326

FIGURE VII
UNIVARIATE TEST SHOWING THE SIGNIFICANT VARIABLE AND POWER

From the univariate test results of sex, we see that the variables which contribute towards the difference in the two groups based on sex are palatal length, post palatal length, zygomatic width, palatal width outside first first upper molars, palatal width inside second upper molars, width between the postglenoid foramen, interorbital width and crown length of first upper molar. As all these eight variables have p value less than 0.05. The only variables which do not contribute is the least width of brain case as the p value is greater than 0.05.

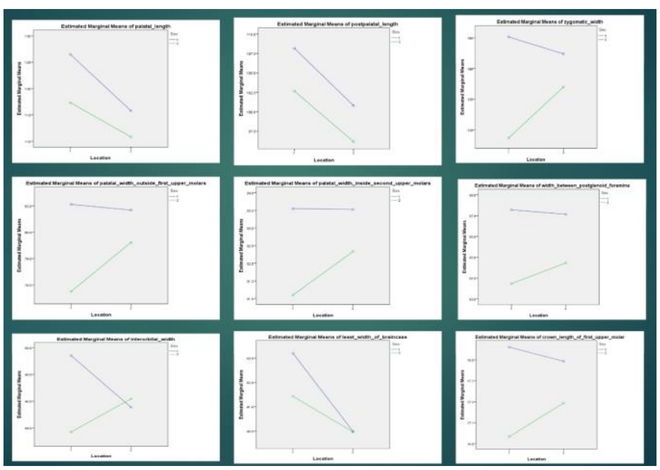

FIGURE VIII
INTERACTION PLOTS

The above figure shows the interaction plots of all the nine morphometric variables. The blue open circles gives the centroids for the males. The green open circles gives the centroids for the females. The length between the blue open circles gives the location effect for males. The length between the green open circles gives the location effect for females. It can be seen that the effect of location is independent of sex only for the case of the palatal length and post palatal length as the two lines seems to be parallel. For all the other seven variables the the effect of location is not independent of sex as the lines are not parallel.

Multivariate Tests of Significance (S = 1, M = 3 1/2, N = 5 1/2)

Test Name	Value	Exact F	Hypoth. DF	Error DF	Sig. of F
Pillais	.94923	27.00549	9.00	13.00	.000
Hotellings	18.69611	27.00549	9.00	13.00	.000
Wilks	.05077	27.00549	9.00	13.00	.000
Roys	.94923				

Note.. F statistics are exact.

Multivariate Effect Size and Observed Power at .0500 Level

TEST NAME	Effect Size	Noncent.	Power
(All)	.94923	243.04940	1.00

Figure IX
EFFECT OF LOCATION - MULTIVARIATE TEST OF SIGNIFICANCE AND POWER

The multivariate statistics of the second main effect location shows that it is statistically significant as the p value of all the three tests is less than 0.05. The power is also observed to be of higher value. So, as the multivariate effect of location is significant we now perform univariate significance test.

EFFECT .. LOCATION (Cont.)
Univariate F-tests with (1,21) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F	ETA Square
palatal_	385.87826	100.60000	385.87826	4.79048	80.55113	.000	.79321
postpala	250.80580	165.93333	250.80580	7.90159	31.74119	.000	.60183
zygomati	40.41884	557.90000	40.41884	26.56667	1.52141	.231	.06755
palata_1	14.17567	89.31767	14.17567	4.25322	3.33292	.082	.13697
palata_2	1.92023	55.90933	1.92023	2.66235	.72125	.405	.03321
width_be	.80719	108.62933	.80719	5.17283	.15604	.697	.00738
interorb	2.53223	159.30067	2.53223	7.58575	.33381	.570	.01565
least_wi	28.52767	138.48900	28.52767	6.59471	4.32584	.050	.17081
crown_le	.10226	5.25267	.10226	.25013	.40884	.529	.01910

Variable	Noncent.	Power
palatal_	80.55113	1.00000
postpala	31.74119	.99971
zygomati	1.52141	.21595
palata_1	3.33292	.41361
palata_2	.72125	.15490
width_be	.15604	.05797
interorb	.33381	.06702
least_wi	4.32584	.50841
crown_le	.40884	.08481

FIGURE X
UNIVARIATE TEST SHOWING THE SIGNIFICANT VARIABLE AND POWER

From the univariate test results of location, we see that the the three variables which contribute towards the difference in the two groups based on location are palatal length, post palatal length and least width of the brain case as the p value is less than 0.05, indicating that it is significant. The other six variables do not contribute towards the difference of the groups as the p value is greater than 0.05, indicating that it is not significant.

From the results of MANOVA, we can conclude that there is significant difference between the two types of species based on the location and sex. But, if we take the interaction effect of the location and sex we can see that it has very low significance.

DISCRIMINANT ANALYSIS

From the analysis of MANOVA, we got to know that there exists substantial differences in the data of rocky mountain wolves and arctic wolves. Hence we further proceed with the discriminant analysis to properly discriminate between the two wolf subspecies and to identify the key variables that helps in differentiating. Discriminant analysis have some assumptions which are same as for MANOVA. All the assumptions holds for the data and hence we can proceed with the discriminant analysis.

Here, we are using location (Rocky mountain wolves - 0 and Arctic wolves - 1) as dependent variables and all the 9 measurement variables as Independent variables. Also, here we have used stepwise discriminant analysis procedure.

location	Rank	Log Determinant
0	4	3.165
1	4	3.367
Pooled within-groups	4	3.882

FIGURE XI
LOG DETERMINANTS

Variables Entered/Removed[a,b,c,d]

		Wilks' Lambda							
						Exact F			
Step	Entered	Statistic	df1	df2	df3	Statistic	df1	df2	Sig.
1	x1	.410	1	1	23.000	33.154	1	23.000	.000
2	x4	.205	2	1	23.000	42.738	2	22.000	.000
3	x9	.114	3	1	23.000	54.555	3	21.000	.000
4	x8	.083	4	1	23.000	55.434	4	20.000	.000

FIGURE XII
VARIABLES ENTERED DURING THE STEPWISE PROCEDURE

The results from the stepwise discriminant analysis shows that the variables x1, x4, x8 and x9 which are palatal length, palatal width outside the first upper molars, least width of brain case and crown length of first upper molars respectively are significant in differentiating the wolves data into two subspecies namely Rocky mountain wolves and Arctic wolves.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	11.087[a]	100.0	100.0	.958

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.083	52.334	4	.000

FIGURE XIII
EIGENVALUES AND SIGNIFICANCE TEST OF FUNCTIONS

The above results shows that one canonical discriminant function is used to discriminate the data. Also, from the Wilk's lambda test we can see that the discriminant canonical function is significant ($p<0.05$).

	Function
	1
x1	2.319
x4	-1.441
x8	.641
x9	-1.371

FIGURE XIV
DISCRIMINANT FUNCTION

The coefficients of the canonical discriminant functions are in the above table.

Classification Results[a,c]

			Predicted Group Membership		
		location	0	1	Total
Original	Count	0	9	0	9
		1	0	16	16
	%	0	100.0	.0	100.0
		1	.0	100.0	100.0
Cross-validated[b]	Count	0	9	0	9
		1	0	16	16
	%	0	100.0	.0	100.0
		1	.0	100.0	100.0

FIGURE XV
CLASSIFICATION TABLE

From the above classification table, we can see that the classification rate is 100 which indicates that all the data are correctly classified into either Rocky mountain wolves or Arctic wolves 100% without any errors what we wanted exactly.

CONCLUSION

In this project we have applied MANOVA and Discriminant analysis statistical methods to the data to check for the relationship between the Rocky mountain wolves and Arctic wolves based on factors like location and sex. From the results, we have concluded that there are substantial differences between Rocky mountain male and female wolves with the Arctic male and female wolves. Also used discriminant analysis to bring up the significant variables that are majorly responsible for the difference along with a discriminant function which help in classifying the wolves based on location.

FUTURE SCOPE

Also we can extend the research by feeding a good amount of data to increase the accuracy of predictions. Also, the other statistical methods like cluster analysis and exploratory factor analysis could be applied to increase the precision of the research.

REFERENCES

[1] "Are Wolves Endangered?" 2011. wolf.org/wow/united-states/are-wolves-endangered/ Accessed: May 1, 2019.

[2] Jolicoeur, P. (1959), "MULTIVARIATE GEOGRAPHICAL VARIATION IN THE WOLF CANIS LUPUS L." Evolution, 13: 283-299. doi:10.1111/j.1558-5646.1959.tb03016.x

[3] "THE STRUCTURE OF THE SKULL IN VERTEBRATES". zoology.ubc.ca/~millen/oldvertebrate/lab6_frameset.htm. Accessed: May 4, 2019.

[4] "Arctic wolf" 2018. en.wikipedia.org/wiki/Arctic_wolf. Accessed May 1, 2019.

[5] "Arctic Wolf" 2019.

worldwildlife.org/species/arctic-wolf Accessed: May 4, 2019.
[6] "Subset selection in regression". psych.colorado.edu/~carey/Courses/PSYC7291/DataSets/Documentation/DetroitDataDoc.txt. Accessed: April 25, 2019.
[7] Shafi, Muhammad A., Rusiman, Mohd S., Hamzah, Nor S. A., et al. 2018. "The analysis of morphometric data on rocky mountain wolves and arctic wolves using statistical method." *IOP Conf. Series:* Journal of Physics: Conf. Series 995 (2018) 012015, doi :10.1088/1742-6596/995/1/012015.
[8] Sharma, S. 1995. "Applied multivariate techniques." S.l.: John Wiley & Sons Limited.

AUTHOR INFORMATION

Sujay Ravi, Graduate student (Master of Science - Industrial and Systems Engineering), Department of Systems Science and Industrial Engineering, Binghamton University.