STA 3180 – Lab #3 Discriminant Analysis

Question 1: Two Group Discriminant Analysis Study: Fitness

Can seven fitness variables discriminate between males and females? The data set "Fitness" contains seven continuous variables measuring aspects of fitness for a group of individuals classified into two categories: males (M) and females (F).

a) Conduct a full two-group discriminant analysis using the data provided. Discuss your findings following the "Discriminant Analysis Guidelines" below.
b) Suppose we have a person of unknown sex with observed values of: Age 50; Weight 75 (in kg); Oxy 45: Runtime 10; RunPulse 170; RstPulse 60; MaxPulse 175. Use your classification routine to predict the sex of this individual.

Discriminant Analysis Guidelines

- Used R code and JMP Output

1. Does a canonical plot indicate discrimination between the groups?
    - The dot plot generated from the R code and JMP output shows that the two groups differ significantly from each other as the inner circles do not intersect.

2. How much of the variation (between group variability) is explained by the discriminant function(s)?
    - The eigenvalue of 1.04828270 is the ratio of the between-groups to the within-groups sums of squares. It further reflects the amount of variance explained by its associated discriminant function.

3. What is the correlation between the discriminant scores (from the covariates) and the grouping variable?
    - The canonical correlation of 0.71539230 is the correlation between the discriminant scores and the grouping variable.

4. What is the significance of the Wilks' Lambda statistic and what does it tell you?
    - The Wilks' Lambda statistic tests the null hypothesis that the means of covariates are equal across all groups. The Wilks' Lambda value of 0.4882139 has a p-value of 0.01142156, making it significant. We therefore reject the null hypothesis that, in the populations from which samples are drawn, there is no difference between the group means on the discriminant function axis.

5. Look at the standardized scoring coefficients (Loadings) and comment on the roles that the variables play in the discrimination.
    - Weight has the largest absolute value (0.899590549), so it is the best discriminator. RunPulse has the lowest absolute value (0.004182857), so it is the worst discriminator.

6. What is the classification accuracy and how does this compare with "chance"?
    - JMP Output
    - The classification accuracy is found to be 26/31 = 0.8387
        - 83.87% classified accuracy
    - $C_{pro} = (\frac{14}{31})^2 + (\frac{17}{31})^2 = 0.50468262226$
    - $C_{pro}\ (1.25) = 0.63085327783$
    - Our classification accuracy of 83.87% is better than chance.

7. Look at the Group Means for all the variables. How do these relate to the discrimination?
    - JMP Output
    - **Discriminant Analysis**

      Group Means

Count	Sex	Age	Weight	Oxy	Runtime	RunPulse	RstPulse	MaxPulse
16	F	47.375000	72.908750	49.83438	10.01625	167.25000	51.31250	171.87500
15	M	48.000000	82.282667	44.75400	11.19400	172.20000	55.73333	175.80000
31	All	47.677419	77.444516	47.37613	10.58613	169.64516	53.45161	173.77419

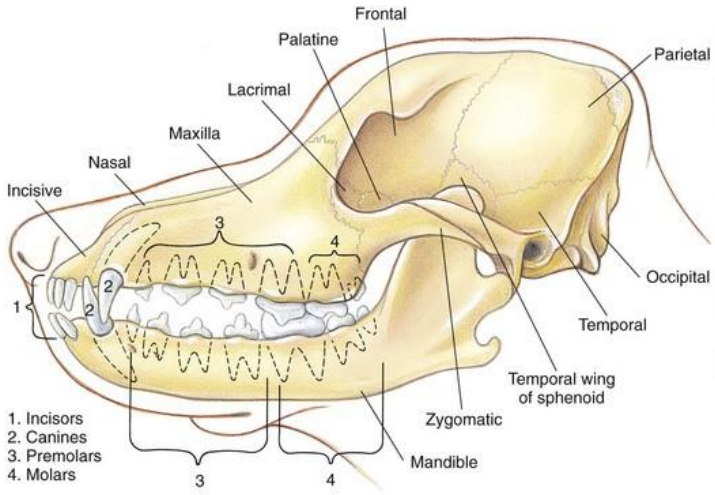
      Discriminant Method: Linear
      Classification: Sex
    - The means of Age and Runtime are very similar, which means that these variables are poor discriminators. Additionally, the difference for RunPulse is substantial, so this contradicts the previous point that it was a bad discriminant. However, Weight has the biggest difference for each Sex confirming the statement that it is a good discriminant.

8. If requested, classify an unknown individual into a group.
    - An unknown individual with the following parameters: Age 50; Weight 75 (in kg); Oxy 45: Runtime 10; RunPulse 170; RstPulse 60; MaxPulse 175
        - The unknown individual is classified as a male.

Question 2: Multi-Group Discriminant Analysis Study: Prehistoric Dogs from Thailand.

Excavations of a prehistoric site in northeast Thailand have produced a collection of canid (dog) bones covering a period from about 3500 BC to the present. However, the origin of the prehistoric dog is not certain. It could descend from the golden jackel (Canis aureus) or the wolf, but the wolf is not native to Thailand. The nearest indigenous sources are western China (Canis lupus chanco) or the Indian subcontinent (Canis lupus pallides).

In order to clarify the ancestors of the prehistoric dogs, mandible (lower jaw) measurements were made on the available specimens. These prehistoric dogs (P) were then compared with the same measurements made on the golden jackal (J), the Indian wolf (IW) and the modern dog (M). See posted data (from B.J.F. Manley).



The nine mandible measurements (in mm) were:

X_1 = length of mandible

X_2 = breadth of mandible below first molar

X_3 = breadth of articular condyle

X_4 = height of mandible below first molar

X_5 = length of first molar

X_6 = breadth of first molar

X_7 = length from first to third molar, inclusive

X_8 = length from first to fourth molar, inclusive

X_9 = breadth of lower canine

(a) Conduct a full discriminant analysis on this data using all variables and canine groups. Follow the discriminant analysis guidelines sheet. Discuss your findings.
Used JMP Output

1. Canonical plot – The inner circles do not intersect, which shows the 4 groups differ significantly from each other.

2. Variation (between-group variability – The eigenvalues of 8.4156 and 3.7237 indicate that the first and second discriminant functions explain a substantial amount of the between-group variability relative to within-group variability. The third function (with an eigenvalue of 0.2233) explains very little variance, so it is likely not as useful for understanding the variance between groups.

3. Correlation between the discriminant scores- The canonical correlation scores indicate the correlation between the discriminant scores and the grouping variable. The canonical correlation scores of 0.9454 and 0.8878 indicate very strong positive correlations for the first 2 models, and the 3rd canonical correlation score of 0.4273 indicates a moderately strong correlation.

4. Significance of the Wilks' Lambda statistic – The Wilks' Lambda statistic tests the null hypothesis that the means of covariates are equal across all groups. The Wilks' Lambda value of 0.0183791 has a p-value of <0.001, making it significant. We therefore reject the null hypothesis that, in the populations from which samples are drawn, there is no difference between the group means on the discriminant function axis.

5. Standardized scoring coefficients – X5 has the highest absolute value at 0.788699 so it is the best discriminator, and X2 has the lowest absolute value at 0.0133337 so it is the worst discriminator.

6. Classification accuracy and Chance – The classification accuracy is found to be 58/60 = 0.967
    - 96.7% classified accuracy
    - $C_{pro} = (\frac{14}{60})^2 + (\frac{20}{60})^2 + (\frac{16}{60})^2 + (\frac{8}{60})^2 = 0.254$
    - C_{pro} (1.25) = 0.318
    - Our classification accuracy of 96.7% is better than chance.

7. Group Means –

Group Means

Count	Group	X1	X2	X3	X4	X5	X6	X7	X8	X9
14	IW	157.35714	11.578571	26.214286	24.714286	24.714286	9.33571	40.21429	44.78571	7.40714
20	J	111.00000	8.180000	18.600000	17.000000	18.200000	6.81500	30.35000	33.35000	4.80500
16	M	125.93750	9.725000	21.375000	21.125000	19.375000	7.67500	32.06250	36.62500	5.86875
10	P	122.80000	10.340000	20.000000	22.900000	19.300000	8.19000	32.80000	35.90000	6.17000
60	All	127.76667	9.745000	21.350000	20.883333	20.216667	7.86167	33.51667	37.31667	5.92333

Discriminant Method: Linear
Classification: Group

    - Based on the group means X2 appears to be a poor discriminator because its means do not differ much for each group. It was said previously that X2 was the worst discriminator based on the standardized scoring coefficients.

(b) A mysterious canine mandible is unearthed at a “dig” in Thailand. Its measurements are in order: 120, 10, 25, 20, 18, 8, 35, 40, 6. Help the archaeologist identify this mysterious canine.

- The predicted Group for the unidentified dog using the model is that it is golden jackal (J).