E-CONTENT PREPARED BY

Dr. Tapajit Bhattacharya

Assistant Professor of Department of

Conservation Biology

Durgapur Government College, Durgapur, West Bengal

(*Affiliated to Kazi Nazrul University, Asansol, West Bengal*)

NAAC Accredited "A" Grade College

(Recognized under Section 2(f) and 12(B) of UGC Act 1956)

E-Content prepared for students of

M.Sc.(Semester-IV) in Conservation Biology

Name of Course:
Biostatistics and Bioinstrumentation

Topic of the E-Content:
Multivariate Analysis

MULTIVARIATE STATISTICS

Habitat data [Compatibility Mode] - Microsoft Excel

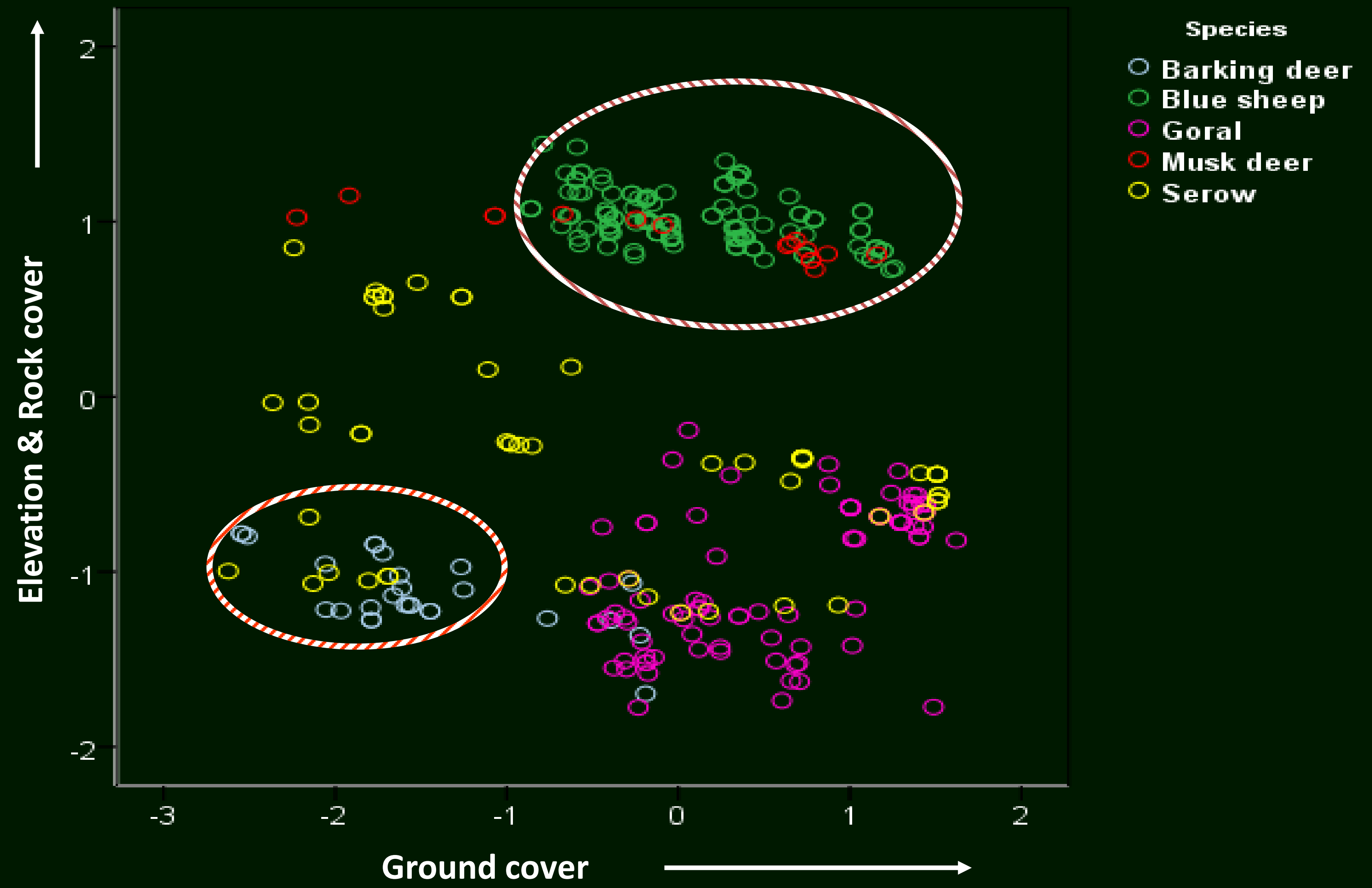
D6 | fx 2

	G	H	I	J	K	L	M	N	O	P	Q	R
1	Slope	Aspect	Altitude(mt)	Grass cover (%)	Grass Height(mm)	Herb Cover(%)	Herb Height(mm)	Habitat type	Lt Depth(mm)	Rock cove	Barren(%)	Cut tree
2	30	NW	3170	0	0	15	2	Oak- Fir- Rh	30	0	85	0
3	45	W	3356	45	35	35	23	Oak- Fir- Rh	50	30	-10	2
4	30	SW	3345	50	60	25	10	Edge	35	0	25	0
5	30	W	3433	60	50	30	15	Edge	20	0	10	0
6	30	W	3433	60	50	30	15	Edge	20	0	10	0
7	30	NW	3360	40	40	50	30	Alpine	5	0	10	0
8	30	NW	3170	0	0	15	8	Oak- Fir- Rh	30	0	85	0
9	15	W	3258	10	40	20	18	Oak- Fir- Rh	35	0	70	1
10	30	W	3450	65	50	35	25	Alpine	0	0	0	0
11	30	W	3433	60	50	30	15	Edge	20	0	10	0
12	15	W	3258	10	40	20	18	Oak- Fir- Rh	35	0	70	1
13	30	NW	3500	60	157	30	35	Alpine	0	10	0	0
14	30	NW	3500	65	140	25	35	Alpine	0	0	10	0
15	30	NW	3500	70	120	20	40	Alpine	0	0	10	0
16	30	NW	3450	65	130	25	45	Alpine	0	10	0	0
17	30	NW	3450	65	75	25	35	Alpine	0	0	10	0
18	30	NW	3450	45	45	30	30	Alpine	0	0	25	0
19	30	NW	3400	55	75	20	30	Alpine	0	0	25	0
20	30	NW	3400	65	120	25	45	Alpine	0	10	0	0
21	30	NW	3400	60	120	30	45	Alpine	0	10	0	0
22	30	NW	3360	60	45	30	40	Alpine	0	0	10	0
23	30	NW	3360	40	40	50	30	Alpine	0	0	10	0
24	30	NW	3360	45	40	50	15	Alpine	0	0	5	0
25	30	NW	3310	45	110	55	15	Edge	10	0	0	0
26	30	NW	3310	40	110	50	20	Edge	15	0	10	0
27	30	NW	3310	40	105	50	25	Edge	30	0	10	2
28	45	NW	3294	45	25	50	15	Edge	20	0	5	2
29	45	NW	3294	30	25	45	10	Edge	34	0	25	0
30	45	NW	3294	35	30	45	15	Edge	25	0	20	0

Sheet2 | Sheet4 | Sheet5 | Monthwise ER of Pellet Group | Sheet6 | Sheet7 | PCA

Ready

Distribution of ungulate evidences along Principal Habitat components in the study area



Ordination Technique

The aim of ordination is twofold

First, it tries to reduce a large number of variables into a smaller number of easier to interpret variables.

And, secondly, it can be used to reveal patterns in multivariate data that would not be identified in univariate analyses

Ordination methods give easy-to-read graphical outputs from relatively easy-to-use software, and this may partly explain their popularity.

Instead of trying to understand variation in a response variable in terms of explanatory variables, in multivariate statistics we look for **structure in the data.**

The problem is that structure is rather easy to find, and all too often it is a feature of that particular data set alone. The real challenge is to find *general structure* that will apply to other data sets as well.

Unfortunately, there is no guaranteed means of detecting pattern, and a great deal of ingenuity has been shown by statisticians in devising means of pattern recognition in multivariate data sets.

The main division is between methods that **assume a given structure** and seek to divide the cases into groups, and methods that seek to **discover structure from inspection of the data frame.**

Different Multivariate techniques:

Principal Component Analysis (PCA)
Factor Analysis
Cluster Analysis
Discriminant Analysis
Neural Networks
Non-metric Multidimensional Scaling (NMDS)
Redundancy Analysis (RDA)
Correspondence Analysis (CA)
Canonical Correspondence Analysis (CCA)

These techniques are not recommended unless you know exactly what you are doing, and exactly why you are doing it.

Beginners are sometimes attracted to multivariate techniques because of the complexity of the output they produce, making the classic mistake of confusing the opaque for the profound.

Principal Component Analysis

The idea of principal component analysis (PCA) is to find a small number of **linear combinations** of the variables so as to capture most of the variation in the data-frame as a whole.

Principal components analysis finds a set of orthogonal standardized linear combinations which together explain all of the variation in the original data.

There are as many principal components as there are variables, but typically it is only the first few that explain important amounts of the total variation.

Calculating principal components is easy. Interpreting what the components mean in scientific terms is hard, and potentially equivocal.

The underlying principle of PCA

Let Y_{ij} be the value of variable j *(j = 1, ..N)* for observation i *(i = 1, .,M). Most* ordination techniques create linear combinations of the variables:

$$Z_{i1} = c_{11}Y_{i1} + c_{12}Y_{i2} + \dots + c_{1N}Y_{iN}$$

The linear combination, $Z_1 = (Z_{11},\dots,Z_{M1})$ *is a vector of length M,* and is called a **principal component**, gradient or axis.

The underlying idea is that the most important features in the N variables are caught by the new variable Z_1

Most ordination techniques are designed in such a way that the first axis is more important than the second, the second more important than the third, etc., and the axes represent different information.

The eigenvalue equation for all axes is given by

(S - IA)C = 0

where C contains the eigenvectors for all axes, I is the identity matrix and A the corresponding eigenvalue

“This may seem like magic to readers not used to matrix algebra, but this expression is fundamental in mathematics and is called the eigenvalue equation for S”---Zuur et al 2007(Analysing Ecological Data)

Eigenvalues in PCA represent the amount of variance explained by each axis.
They can be expressed as numbers, percentage of the total variance, or as cumulative percentage of the total variance

Table 12.2. Eigenvalues and eigenvalues expressed as cumulative percentage. Some software packages rescale the eigenvalues so that the sum of all eigenvalues is equal to 1. These are given in the second column. Unscaled eigenvalues are in the third column. The sum of all eigenvalues is 7.

Axis	Eigenvalue (scaled)	Eigenvalue (unscaled)	Cumulative Eigenvalue as %
1	0.647	4.529	64.703
2	0.163	1.144	81.045
3	0.063	0.440	87.326
4	0.049	0.342	92.206