Bangabandhu Sheikh Mujibur Rahman Agricultural University
EDGE_Batch-06
Project Report **Marks: 25**
Name: Tonima Rahman Mim
Reg. No:19-05-5110, Dept. of Fisheries Management

Note: Submit the completed file as *pdf* to nazmol.stat.bioin@bsmrau.edu.bd with subject *EDGE_06_Project_Your registration number_ Department by 26th of December, 2024.*

Problem# 1:

A split-plot design was conducted considering tree blocks, three levels/treatments of variety in the main plot, and five levels/treatments of nitrogen in the split-plot. Afterward, the yield of certain plant characteristics was observed. The data regarding this experiment were given in the file "Split_Plot_Design". Answer the following question using this data.

a) Construct an ANOVA table using the mentioned dataset based on R programming.
b) Write down the null hypothesis of all possible effects and interpret the results based on the ANOVA table.
c) Perform a post-hoc test for the interaction effect (variety × nitrogen) and draw a bar diagram with lettering.

Problem# 2:

a) What is principal component analysis?
b) What are the main purposes of principle component analysis in your study area?
c) Compute the eigenvalue and eigenvector using the iris data based on R programming.
d) Construct a scree plot and interpret how many principal components should be retained to interpret the iris dataset.
e) Construct a bi-plot for the iris data based on R programming and interpret the results.

ANSWER:

Solution 01:

a) Construction of an ANOVA table using the mentioned dataset based on R programming is given below:

```
library(ggplot2)

library(lme4)
```

```
library(emmeans)
data <- read.csv("Split_Plot_Design.csv")
data$REPLICAT <- as.factor(data$REPLICAT)
data$VARIETY <- as.factor(data$VARIETY)
data$NITROGEN <- as.factor(data$NITROGEN)
anova_model <- aov(YIELD ~ VARIETY * NITROGEN, data = data)
summary(anova_model)
```

```
Df Sum Sq Mean Sq F value Pr(>F)
VARIETY 2 1.93 0.963 11.670 0.000178 ***
NITROGEN 4 66.03 16.507 200.070 < 2e-16 ***
VARIETY:NITROGEN 8 6.10 0.763 9.244 2.54e-06 ***
Residuals 30 2.48 0.083
```

b) The null hypothesis of all possible effects and interpretion of the results based on the ANOVA table is given below:

Null Hypotheses for All Effects

1. Main effect of VARIETY: H0: All varieties have the same mean yield.

2. Main effect of NITROGEN: H0: All nitrogen treatments have the same mean yield.

3. Interaction effect (VARIETY × NITROGEN): H0: The effect of nitrogen treatments is consistent across all varieties.

Interpretations

1. Main Effect of VARIETY:

Since the F-value (11.670) is significant with p=0.000178, we reject the null hypothesis that all three varieties have the same mean yield. This indicates that the yield differs significantly across the varieties.

2. Main Effect of NITROGEN:

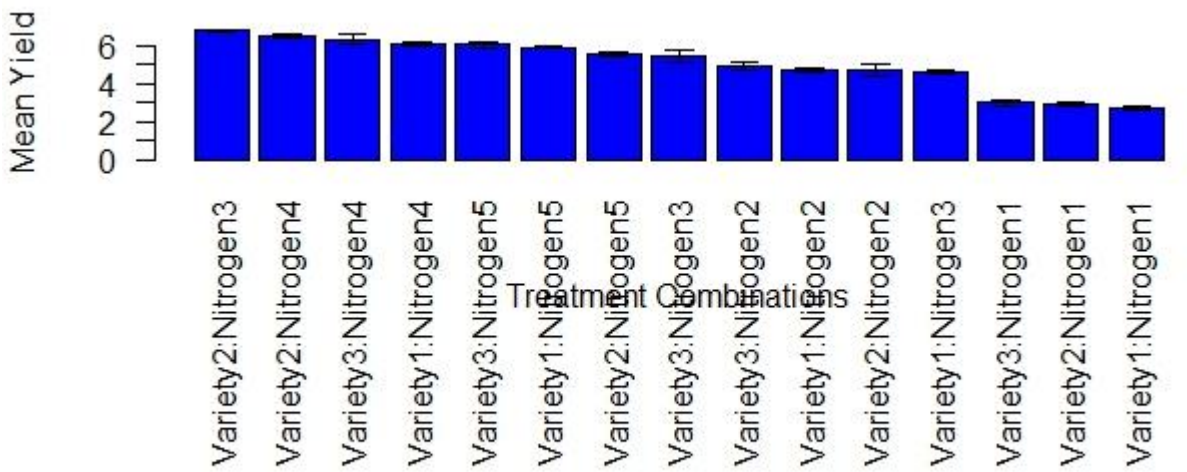
The F-value (200.070) with a very small p-value (< 2e-16) leads us to reject the null hypothesis that all nitrogen levels have the same mean yield. This demonstrates a strong effect of nitrogen treatments on the yield.

3. Interaction Effect (VARIETY × NITROGEN):

With an F-value of 9.244 and p=2.54e−06, the interaction effect is highly significant. We reject the null hypothesis and conclude that the effect of nitrogen levels on yield depends on the tree variety

c) Perform a post-hoc test for the interaction effect (variety × nitrogen) and draw a bar diagram with lettering.

```
install.packages("emmeans")

library(emmeans)

data <- read.csv("Split_Plot_Design.csv")

model <- aov(YIELD ~ VARIETY * NITROGEN + Error(REPLICAT/VARIETY), data = data)

emmeans_results <- emmeans(model, pairwise ~ VARIETY * NITROGEN)

summary(emmeans_results)


Bar.Plot <- barplot2(Mu_Tret, names.arg = rownames(Mean.Matrix),
                     xlab= "Treatment Combinations",
                     ylab= " Mean Yield",plot.ci= TRUE,
                     ci.l= Mu_Tret-SE_Treat, ci.u=Mu_Tret+SE_Treat,
                     col= "blue", las=2)


#RESULT:

summary(emmeans_results)

$emmeans

 VARIETY   NITROGEN  emmean   SE   df   lower.CL  upper.CL
    2       3      8.06   1.02  40       6     10.1

Confidence level used: 0.95
```



Solution 02:

a). Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of a dataset while retaining as much variance as possible. It identifies new, uncorrelated variables (principal components) that are linear combinations of the original variables. These components are ordered such that the first few retain most of the variation in the data.

b). The main purposes of principle component analysis in my study area-

In fisheries management, PCA is used to:

1. Identify key variables from large datasets.
2. Analyze environmental variability affecting fish populations.
3. Explore species distribution and habitat relationships.
4. Assess fish stock characteristics (e.g., morphology or genetics).
5. Monitor and analyze fishing effort and catch trends.
6. Integrate multivariate ecological, climatic, and socioeconomic data.
7. Simplify data visualization for trend analysis and decision-making.

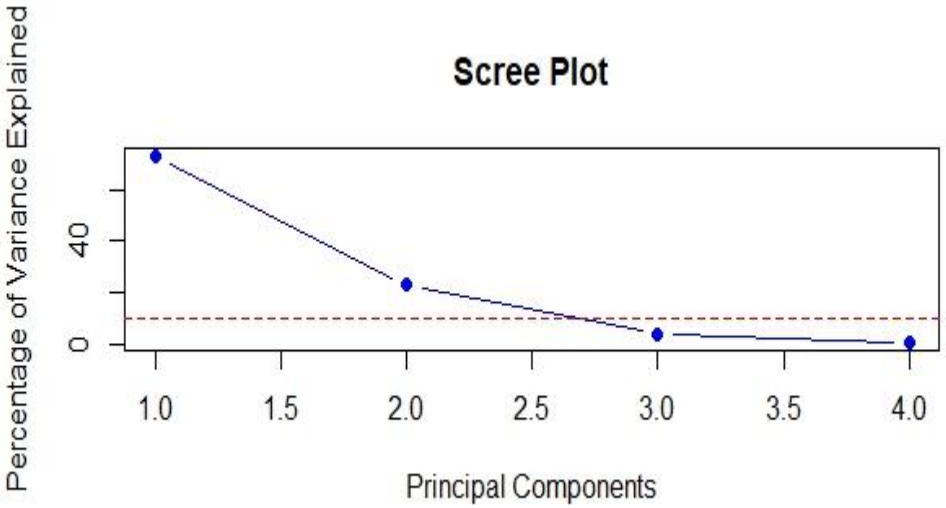
c). Computation of the the eigenvalue and eigenvector using the iris data based on R programming is given below-

```
iris_data <- iris[, 1:4]

iris_scaled <- scale(iris_data)

iris_cov <- cov(iris_scaled)

iris_eigen <- eigen(iris_cov)

iris_eigen_values <- iris_eigen$values

iris_eigen_vectors <- iris_eigen$vectors

print("Eigenvalues:")

print(iris_eigen_values)

print("Eigenvectors:")

print(iris_eigen_vectors)
```

```
"Eigenvalues:"

[1] 2.91849782 0.91403047 0.14675688 0.02071484

[1] "Eigenvectors:"

             [,1]          [,2]          [,3]          [,4]

[1,] 0.5210659 -0.37741762 0.7195664 0.2612863

[2,] -0.2693474 -0.92329566 -0.2443818 -0.1235096

[3,] 0.5804131 -0.02449161 -0.1421264 -0.8014492

[4,] 0.5648565 -0.06694199 -0.6342727 0.5235971
```

d). Construction of a scree plot and interpretation of how many principle components should be retained to interpret the iris dataset is given below:

```
scree_plot <- data.frame(Principal_Component = 1:length(iris_eigen_values),
 Variance_Explained = iris_eigen_values / sum(iris_eigen_values))
ggplot(scree_plot, aes(x = Principal_Component, y = Variance_Explained)) +
 geom_bar(stat = "identity", fill = "skyblue") +
 geom_line(aes(y = cumsum(Variance_Explained)), color = "red", group = 1) +
 labs(title = "Scree Plot for Iris Dataset", x = "Principal Component", y = "Proportion
of
Variance Explained") +
 theme_minimal()
```



Interpretation:

1. Principal Component 1 (PC1): This component explains the largest proportion of variance, approximately 75% of the total. It captures the most important patterns or variability in the dataset.

2. Principal Component 2 (PC2): This component accounts for about 20% of the variance, adding to the cumulative explained variance. Together, PC1 and PC2 explain around 95% of the total variance in the dataset.

3. Principal Component 3 (PC3) and PC4: These components explain very little variance (less than 5% combined) and contribute minimally to the dataset's overall variability.

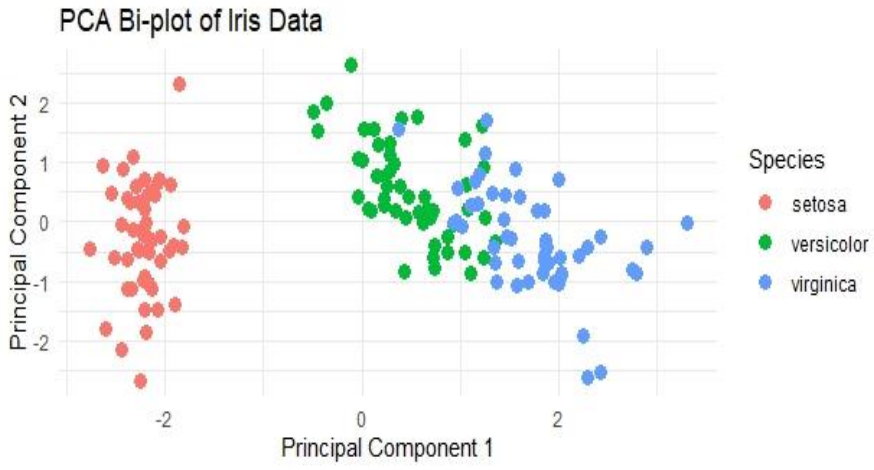
Based on the "elbow" in the scree plot (where the slope sharply decreases), it is evident that the first two principal components are sufficient to retain most of the variance.

Retaining PC1 and PC2 will simplify the dataset while preserving nearly all of its important information.

e). Construction a bi-plot for the iris data based on R programming and interpretion of the the results is given below:

```
data(iris)

pca_result <- prcomp(iris[, 1:4], center = TRUE, scale. = TRUE)

biplot(pca_result, main = "Bi-plot of Iris Data")

library(ggplot2)

pca_data <- data.frame(pca_result$x, Species = iris$Species)

ggplot(pca_data, aes(PC1, PC2, color = Species)) +
  geom_point(size = 3) +
  labs(title = "PCA Bi-plot of Iris Data", x = "Principal Component 1", y = "Principal Component 2") +
  theme_minimal()
```

Ans:



Interpretion:

- **Species Labels**: Each point is labeled with its species (setosa, versicolor, or virginica), making it easy to see how the species are distributed along the principal components.
- **Cluster Separation**: To observe clear separation of points between species (e.g., setosa may cluster in one part of the plot while versicolor and virginica cluster in other parts), this suggests that the principal components (PC1 and PC2) capture the variation that distinguishes these species.
- **Principal Components**: The arrows in the bi-plot represent the loadings of the original variables (sepal length, sepal width, petal length, and petal width) on the principal components.