

Bangabandhu Sheikh Mujibur Rahman Agricultural University

EDGE_Batch-06

Project Report Marks: 25

Name:Fariha Khan Mahi.....

Reg. No:.....18-05-4558.....Dept.....Entomology.....

Note: Submit the completed file as pdf to nazmol.stat.bioin@bsmrau.edu.bd and rabiulauwul@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.

```
summary(model)

Error: REPLICAT
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  2  1.255   0.6274

Error: REPLICAT:VARIETY
      Df Sum Sq Mean Sq F value Pr(>F)
VARIETY  2  1.9256   0.9628  24.58 0.00566 **
Residuals  4  0.1567   0.0392
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: within
      Df Sum Sq Mean Sq F value Pr(>F)
NITROGEN  4  66.03  16.507  372.43 < 2e-16 ***
VARIETY:NITROGEN  8   6.10   0.763  17.21 3.35e-08 ***
Residuals  24   1.06   0.044
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
```

- b) Write down the null hypothesis of all possible effects and interpret the results based on the ANOVA table.

Null Hypotheses for All Possible Effects

Based on the split-plot design and the given ANOVA table, the null hypotheses for each effect are:

Main Plot Effect (VARIETY):

Null Hypothesis (H_0): The means of yield are the same across all levels of VARIETY.

$H_0: \mu_1 = \mu_2 = \mu_3$

Alternative Hypothesis (H_1): At least one level of VARIETY has a different mean yield.

Split-Plot Effect (NITROGEN):

Null Hypothesis (H_0): The means of yield are the same across all levels of NITROGEN.

$H_0: \mu_{N1} = \mu_{N2} = \mu_{N3} = \mu_{N4} = \mu_{N5}$

Alternative Hypothesis (H_1): At least one level of NITROGEN has a different mean yield.

Interaction Effect (VARIETY:NITROGEN):

Null Hypothesis (H_0): There is no interaction between VARIETY and NITROGEN on yield.

H_0 : The effect of NITROGEN is independent of VARIETY.

Alternative Hypothesis (H_1): There is an interaction between VARIETY and NITROGEN on yield.

Interpretation of the ANOVA Table

Main Plot Effect (VARIETY):

P-value: 0.00566 (significant at the 0.01 level).

Conclusion: Reject H_0 . The yields are significantly different across the three levels of VARIETY. This indicates that VARIETY has a significant effect on yield.

Split-Plot Effect (NITROGEN):

P-value: $< 2e-16$ (extremely significant).

Conclusion: Reject H_0 . The yields are significantly different across the five levels of NITROGEN. This indicates that NITROGEN has a highly significant effect on yield.

Interaction Effect (VARIETY:NITROGEN):

P-value: 3.35e-08 (extremely significant).

Conclusion: Reject H_0 . There is a significant interaction between VARIETY and NITROGEN. This means the effect of NITROGEN on yield depends on the level of VARIETY.

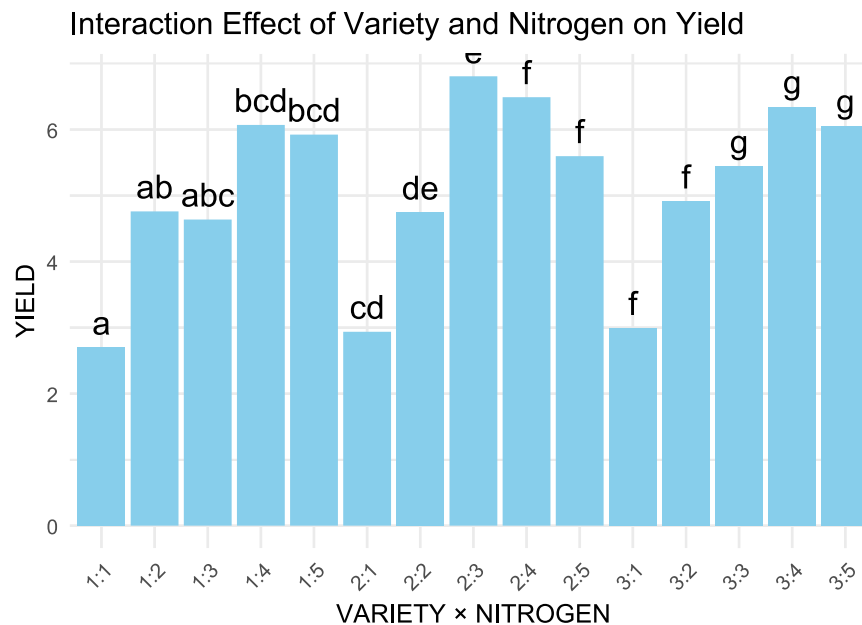
So ,

Both VARIETY and NITROGEN have significant main effects on yield.

There is a significant interaction between VARIETY and NITROGEN, suggesting that the influence of nitrogen levels on yield varies across different varieties.

Further post-hoc analysis (e.g., Tukey's HSD test) can be conducted to identify which levels are significantly different if needed. Let me know if you want help with that!

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



Problem# 2:

- a) What is principal component analysis?

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of a dataset while preserving as much variability as possible. It transforms the original set of correlated variables into a smaller set of uncorrelated variables called **principal components**.

b) What are the main purposes of principle component analysis in your study area?

In entomology, **Principal Component Analysis (PCA)** serves several important purposes by simplifying complex datasets and uncovering patterns that help researchers study insects more effectively. The main applications include:

1. Morphological Studies

PCA is widely used to analyze and compare the morphology of insects by reducing complex measurements into principal components that represent the most significant variations. Differentiating insect species or populations based on wing size, shape, or antennal structures.

2. Species and Population Differentiation

PCA helps identify patterns in traits that distinguish species, subspecies, or insect populations.

Example: Comparing populations of pest species to identify regional adaptations or ecological preferences.

3. Behavioral Studies

Analyzing complex behavioral data using PCA helps entomologists identify the key factors that influence insect behavior.

Example: Studying the responses of insects to various stimuli (e.g., light, chemicals) to uncover significant patterns.

4. Environmental and Ecological Analysis

PCA is used to evaluate the relationship between insect populations and environmental variables (e.g., temperature, humidity, vegetation).

Example: Identifying which environmental factors are most influential in determining the distribution of a pest species.

5. Chemical Ecology and Semiochemical Analysis

PCA is applied in studies involving the chemical compounds used by insects for communication or defense.

like Grouping volatile compounds to identify key differences between pheromones of different insect species.

6. Genetics and Genomics

In entomological genetics, PCA is used to analyze genetic variation or gene expression data.

For example: Studying the genetic diversity of pest populations to inform control strategies.

7. Pest Management

PCA helps in Integrated Pest Management (IPM) by identifying the key factors that influence pest prevalence or pesticide resistance. Example: Determining which environmental, genetic, or chemical factors contribute to resistance in *Tuta absoluta* or *Plutella xylostella*.

8. Epidemiology and Vector Studies

PCA is used in the study of insect vectors to identify key traits or environmental factors associated with their ability to transmit diseases.

Example: Analyzing the role of environmental conditions in the distribution of mosquitoes that transmit malaria or dengue.

- c) Compute the eigenvalue and eigenvector using the iris data based on R programming.

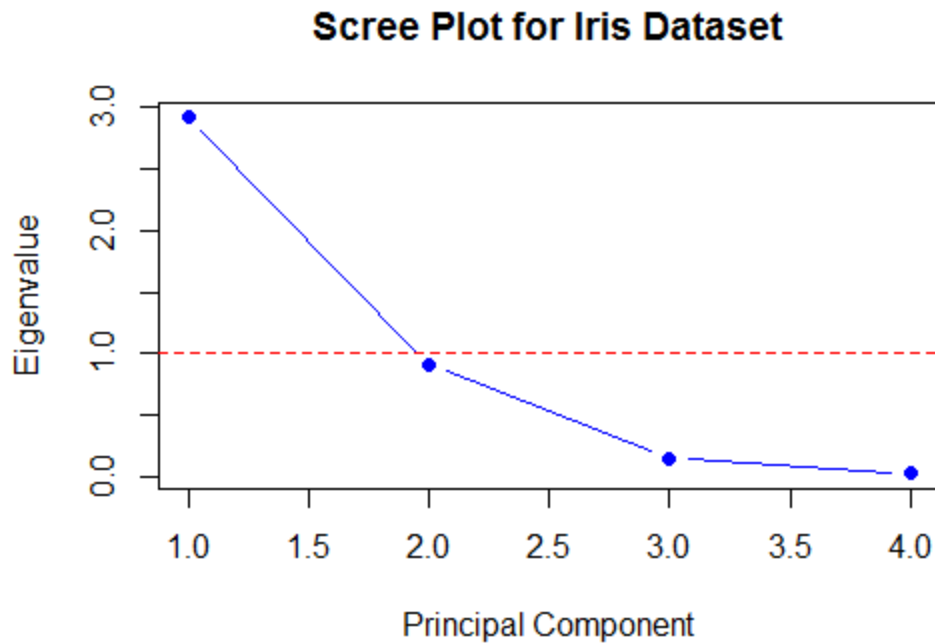
Eigenvalues:

```
> print(eigen_result$values)
[1] 2.91849782 0.91403047 0.14675688 0.02071484
> cat("\nEigenvectors:\n")
```

Eigenvectors:

```
> print(eigen_result$vectors)
      [,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) Construct a scree plot and interpret how many principle components should be retained to interpret the iris dataset.

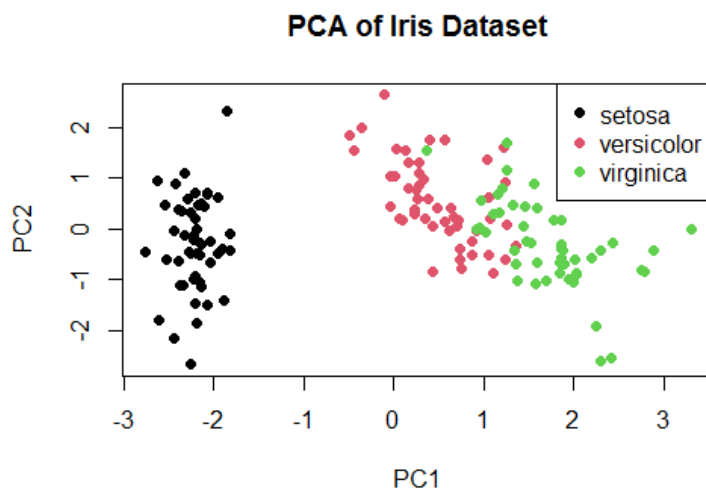
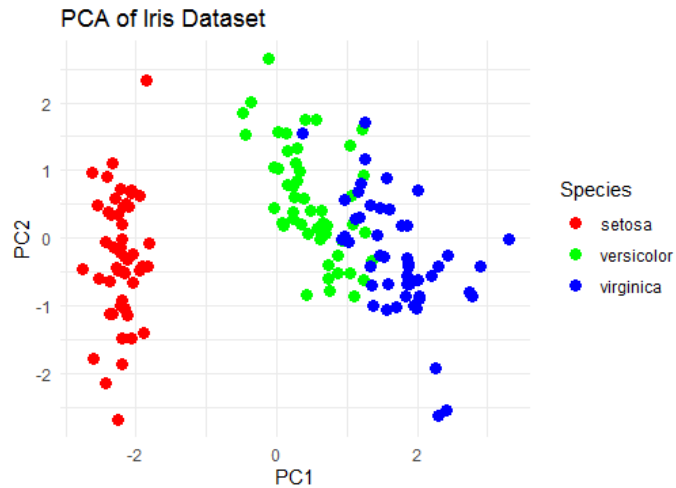


1. **X-axis (Principal Components):** This axis lists the principal components in order of their significance, from PC1 to PC4.
2. **Y-axis (Eigenvalues):** This axis shows the eigenvalues corresponding to each principal component. The eigenvalues represent the amount of variance explained by each component.
3. **Elbow Point:** The "elbow" of the scree plot is the point where the curve starts to flatten out. This is usually where the eigenvalues start to drop off significantly.

Analysis:

- **PC1:** The first principal component has an eigenvalue significantly higher than 1, indicating it captures a substantial amount of the variance in the dataset.
- **PC2:** The second principal component also has an eigenvalue above 1, though it is less than that of PC1.
- **PC3 and PC4:** Both of these components have eigenvalues below 1, indicating they contribute less to the overall variance.

e) Construct a bi-plot for the iris data based on R programming and interpret the results.



- **Axes and Variance:**

- The x-axis represents the first principal component (PC1).
- The y-axis represents the second principal component (PC2).

- **Data Points:**

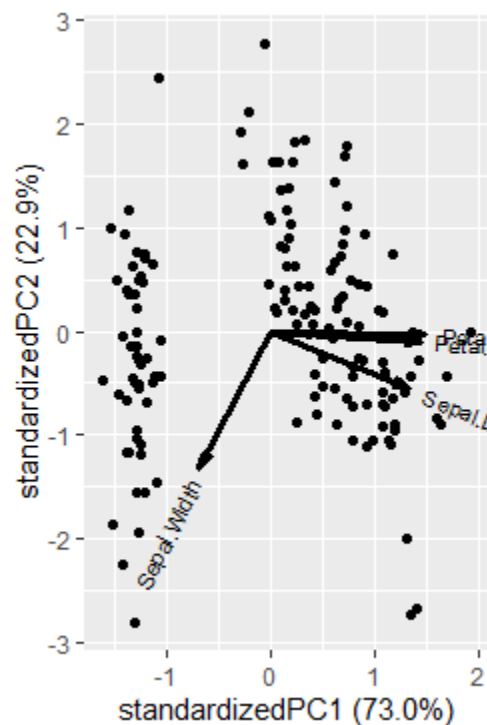
- Each point on the plot represents an individual iris flower from the dataset.
- The data points are color-coded based on the species of the iris flowers:
 - Black dots represent the **Setosa** species.
 - Red dots represent the **Versicolor** species.
 - Green dots represent the **Virginica** species.

- **Separation of Species:**

- The **Setosa** species is well-separated from the other two species along the first principal component (PC1).
- **Versicolor** and **Virginica** show some overlap but are still distinguishable, primarily along the second principal component (PC2).

- **Principal Components:**

- PC1 captures the maximum variance in the data, indicating that it is the most significant direction of data variability.
- PC2 captures the second-highest variance, orthogonal to PC1



- The biplot helps in understanding the relationships between the variables and the principal components.
- It shows that Petal.Length and Petal.Width have a strong influence on PC1, while Sepal.Width has a stronger influence on PC2.
- The spread of data points along the vectors can help in identifying clusters or patterns within the iris dataset.