BELLS UNIVERSITY OF TECHNOLOGY

COURSE CODE: ICT 313

COURSE TITLE: IBM SPSS LEVEL 1

COLLEGE: COLLEGE OF NATURAL AND APPLIED SCIENCES (COLNAS)

DEPARTMENT: BIOLOGICAL SCIENCES

PROGRAMME: BIOTECHNOLOGY

GROUP NUMBER: GROUP 2

TEAM MEMBERS:

1. Lawal Khalid Adewunmi **2022/11826**
2. Madu Angel Amarachi **2023/12789**
3. Bowade Ifeoluwa Christiana **2022/11601**
4. Abudu Kanyinsola Inumidun **2022/11640**
5. Wanikone Tolulope Mary **2022/11362**
6. Aledare Temidun Enitan **2022/11130**

TABLE OF CONTENTS

1. PROJECT SUMMARY

All groups were required to choose a topic from the predefined section relevant to IBM SPSS Statistics Application. The topic was chosen based on departmental strengths. The group is also required to produce a report in line with academic standards, referencing similar scholarly articles, journals to inform your work is encouraged. All implemented projects must then be published on GitHub with each member added as a contributor.

---

2. INTRODUCTION ; BIODIVERSITY AND WHY WE CHOSE IT

Biodiversity is a critical measure of the health of ecosystems reflecting the variety of organisms within a particular environment. As biotechnology students, understanding factors that influence and affect organisms and their environment is essential for conversation efforts and sustainable ecosystem management.
This report presents a comprehensive analysis of a biodiversity dataset. This analysis uses statistical techniques including : regression analysis, factor analysis and ANOVA to discover patterns and relationships within the data.

---

3. OBJECTIVES

The primary objectives of this analysis are ;

- To determine the relationship between environmental variables and biodiversity indices.
- To identify underlying factors that influence biodiversity.
- To provide insights for biodiversity conservation and ecosystem management.
- To check for significant differences in biodiversity indices across different ecosystem types.

---

4. METHODOLOGY

The analysis was conducted using IBM SPSS which is a statistical software used for data analysis. The following techniques were used;

- Regression Analysis; this was used to examine the relationship between biodiversity indices and environmental variables like soil type, rainfall,etc.
- ANOVA (Analysis of Variance); this was used to test for significant differences in biodiversity indices across various ecosystem types.
- Factor Analysis; this was used to identify underlying factors that explain variance in the data set.

The dataset used contains 300 observations with variables including ecosystem type, location, size of area surveyed,dominant species, latitude/longitude, vegetation type, etc.

---

5. DATA DESCRIPTION

The data set comprises 300 entries, each representing a unique ecosystem. Some key variables include;

- Ecosystem type; forest, marine, wetland. Desert, grassland, savanna
- Location; geographic location of the ecosystem.
- Size of area surveyed
- Sampling effort; time taken to survey the ecosystem.
- Biome type; tropical, temperate and polar
- latitude/longitude; geographic coordinates of ecosystem
- Animal migration routes; presence/absent
- Primary economic activity; the main economic activity in the ecosystem e.g fishing, tourism, etc.
- Ecosystem age
- Vegetation type; rainforest, grassland, tundra.
- Dominant species
- Threat level; the level of threat to the ecosystem
- Average annual rainfall
- Biodiversity index; measure of biodiversity in the ecosystem.
- Soil type; loamy, clay,sandy,peat.
- Carbon sequestration potential; the potential of the ecosystem to sequester carbon.

---

6. **Analysis and Results**

6.1 **Regression Analysis**

This is a statistical method that was used to study the relationship between the dependent variables (Biodiversity index) and some environmental factors some of which are: Temperature , Annual Rainfall, soil type,e.t.c. These environmental factors are the independent variables.

- Analysis Summary:
    - o R-squared(shows how much of the change in biodiversity index is explained by temperature, soil type, e.t.c.): 0.022
    - o Adjusted R-squared: 0.002
    - o Standard Error of the Estimate: 14.650

The low R-squared value shows that the analysis explains only 2.2% of Biodiversity changes are affected by the environmental factors. This goes to say that these environmental factors do not significantly explain changes in the Biodiversity.

- ANOVA Results:
    - o F-statistic: 1.111
    - o p-value: 0.356

The p-value tests if the relationship is important and if the value is < 0.05, the relationship is important or better said as **“statistically significant”.** From our result, we see that the p-value is far greater than 0.05, indicating that the relationship is not important or statistically significant

- Coefficients(shows how much the Biodiversity index changes with a change by 1 unit in the environmental factors):
    - o None of the environmental factors showed meaningful effects on the Biodiversity Index.

The regression analysis carried out tells us the environmental factors used for this particular analysis do not have any major impact on the Biodiversity Index in this dataset. It could mean that other environmental factors may play a more visible role in determining biodiversity.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.149[a]	.022	.002	14.650

a. Predictors: (Constant), SoilType_Code=5.0, Ecosystem Age (years), Average Annual Rainfall (mm), SoilType_Code=4.0, SoilType_Code=2.0, SoilType_Code=3.0

6.2 **ANOVA Analysis**

ANOVA(Analysis of Variance) is of two types: One-way ANOVA and Two-way ANOVA. For the purpose of this analysis, the One-way ANOVA was employed and it was used to test for meaningful differences in Biodiversity Index over various types of ecosystems.

- ANOVA Results:
    - o F-statistic(shows whether there are differences between ecosystems): 0.182

- o P-value: 0.969

The p-value is much greater than 0.05, telling us that the relationship between Biodiversity Index over ecosystems is not statistically significant.

- · Post Hoc Tests(Tukey HSD)
  - o No important differences were found between pairs of ecosystem types.

The ANOVA results imply that biodiversity is rather consistent over various ecosystems due to the insignificant differences over ecosystems as seen from the results.

ANOVA

Biodiversity Index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	197.932	5	39.586	.182	.969
Within Groups	64117.815	294	218.088		
Total	64315.747	299			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Biodiversity Index	Eta-squared	.003	.000	.002
	Epsilon-squared	-.014	-.017	-.015
	Omega-squared Fixed-effect	-.014	-.017	-.015
	Omega-squared Random-effect	-.003	-.003	-.003

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Tukey HSD[a,b]

Ecosystem Type Recoded	N	Subset for alpha = 0.05 1
6.00	63	75.83
5.00	49	76.18
3.00	49	76.24
2.00	48	76.96
4.00	35	77.86
1.00	56	77.88
Sig.		.984

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 48.406.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

6.3 **Factor Analysis**

This is a statistical method used to identify patterns or relationships among Environmental Factors/variables by treating them as underlying factors explaining the dataset.

- KMO and Bartlett's Test:
  - KMO Measure of Sampling Adequacy: 0.561 (borderline acceptable)
  - Bartlett's Test of Sphericity: p-value = 0.593 (insignificant)

The KMO value has a recommended threshold of 0.6 and we can see that our result is below that threshold. Bartlett's test is not significant, implying that the data may not be the best pick for factor analysis.

- Total Variance Explained:
  - Two components were drawn, interpreting approximately 38% of the total variance.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Average Annual Rainfall (mm)	1921.2304	1129.28508	300
Ecosystem Age (years)	4668.54	2841.232	300
Biodiversity Index	76.75	14.666	300
Carbon Sequestration Potential (tonnes/year)	5015.0480	2820.38650	300
Size of Area surveyed (sq km)	270.1492	137.60230	300
Sampling Effort (hours)	56.09	26.958	300

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.561
Bartlett's Test of Sphericity	Approx. Chi-Square	13.115
	df	15
	Sig.	.593

Communalities

	Initial	Extraction
Average Annual Rainfall (mm)	1.000	.432
Ecosystem Age (years)	1.000	.383
Biodiversity Index	1.000	.312
Carbon Sequestration Potential (tonnes/year)	1.000	.474
Size of Area surveyed (sq km)	1.000	.360
Sampling Effort (hours)	1.000	.318

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.256	20.940	20.940	1.256	20.940	20.940	1.252	20.874	20.874
2	1.023	17.056	37.996	1.023	17.056	37.996	1.027	17.122	37.996
3	.992	16.536	54.533						
4	.955	15.914	70.447						
5	.913	15.209	85.656						
6	.861	14.344	100.000						

Extraction Method: Principal Component Analysis.

7. **Discussion**

From the Analysis and Results we can confidently say that:

1. The Regression Analysis shows that selected Environmental Factors(temperature, soil type, annual rainfall,e.t.c) do not clearly or precisely explain the differences in Biodiversity Index. This could be a function of Biodiversity Dynamics affected by a wide range of other factors not included in this dataset such as species interaction and external activities.

2. The ANOVA results suggest that the Biodiversity Index has no statistical significance over various ecosystem types, which could mean that biodiversity is rather consistent over various ecosystems.

3. The Factor Analysis, while not statistically sturdy, identified potential underlying factors in relation to productivities of ecosystems and intensities of sampling. These communicate to us that upcoming research should take into account additional variables and a more sturdy statistical method for better understanding the factors affecting Biodiversity Index.

---

8 . **Conclusion.**

The analysis gives valuable information about the Environmental Factors in relation to the Biodiversity and its Index. While the Regression and ANOVA did not bring to light or make obviously known the actual important relationships, the factor analysis showed possible underlying factors that require more study and research. The results pinpoint that Biodiversity is somewhat complex and there is a need for more extensive and detailed data collection and analysis in further studies to come.