Exploring Biodiversity Trends Using Proportional Species Richness

Sulaiman Dauda

2023-04-02

Data Exploration

The data consists of seven variables; Bird, Carabids, Hoverflies, Isopods, Macromoths, Grasshoppers & Crickets, and Vascular plants. These variables represent proportional species richness, which is the number of different species in each taxonomy. In this section, we will explore these seven variables as well as the correlations between them.

Univariate Analysis: The table below shows the mean, standard deviation, and skewness for each of the seven variables.

Table 1: Variable Summary Statistics

taxi_group	mean	sd	skewness
Hoverflies	0.58	0.06	0.12
Macromoths	0.83	0.07	0.13
Vascular_plants	0.76	0.08	0.3
Grasshoppers_._Crickets	0.54	0.08	0.31
Bird	0.9	0.08	0.32
Isopods	0.63	0.16	-0.25
Carabids	0.64	0.16	-1.08

From the table, it is evident that the mean values for each variable range from 0.54 to 0.9. The standard deviation for each variable ranges from 0.06 to 0.16. The skewness values range from -1.08 to 0.32.

It is interesting to note that the skewness value for the "Carabids" variable is -1.08, which indicates that the distribution is negatively skewed, i.e., the tail of the distribution is longer on the left-hand side. In contrast, the skewness value for the "Hoverflies" variable is 0.12, indicating a slightly positively skewed distribution.

Correlation Analysis: The correlation matrix below shows the pairwise correlation coefficients between the seven variables.

Table 2: Correlation Matrix

	Bird	Carabids	Hoverflies	Isopods	Macromoths	Grasshoppers_._Crickets	Vascular_plants	Easting	Northing
Bird	1.00	0.45	0.36	-0.07	0.83	0.76	0.44	0.49	-0.16
Carabids	0.45	1.00	0.62	0.74	0.13	0.22	0.71	0.19	-0.32
Hoverflies	0.36	0.62	1.00	0.55	-0.03	0.02	0.71	0.21	-0.05

	Bird	Carabids	Hoverflies	Isopods	Macromoths	Grasshoppers_._Crickets	Vascular_plants	Easting	Northing
Isopods	-0.07	0.74	0.55	1.00	-0.38	-0.30	0.65	0.01	-0.32
Macromoths	0.83	0.13	-0.03	-0.38	1.00	0.80	0.07	0.22	-0.33
Grasshoppers_._Crickets	0.76	0.22	0.02	-0.30	0.80	1.00	0.12	0.22	-0.30
Vascular_plants	0.44	0.71	0.71	0.65	0.07	0.12	1.00	0.40	-0.24
Easting	0.49	0.19	0.21	0.01	0.22	0.22	0.40	1.00	0.35
Northing	-0.16	-0.32	-0.05	-0.32	-0.33	-0.30	-0.24	0.35	1.00

The correlation matrix shows that the strongest positive correlations are between Bird and Macromoths (0.83), Bird and Grasshoppers & Crickets (0.76), and Hoverflies and Carabids (0.62). The strongest negative correlations are between Isopods and Macromoths (-0.38), Isopods and Northing (-0.32), and Grasshoppers & Crickets and Northing (-0.30). The variable with the weakest correlation to the other variables is Easting.

Looking at the R2 values (squared correlation coefficients), we see that the variables with the highest correlation to the other variables are Bird (0.67 average R2) and Macromoths (0.42 average R2). In contrast, Easting and Northing have the lowest average R2 values (0.01 and 0.05, respectively).

Hypothesis tests

One-sample t-test: This test can be used to determine whether the mean of a sample is significantly different from a hypothesized population mean. In this case, we can use a one-sample t-test to determine whether the mean BD7 change in the study area is significantly different from zero.

	t	df	p_value
t	-9.314601	43	0

The result of the one-sample t-test shows that the t-value is -9.3146 with a degrees of freedom of 43 and a p-value of 7.118e-12. Since the p-value is much smaller than the significance level of 0.05, we can reject the null hypothesis that the mean BD7 change is equal to zero. Therefore, we can conclude that there is a significant change in the biodiversity of the study area between the two periods.

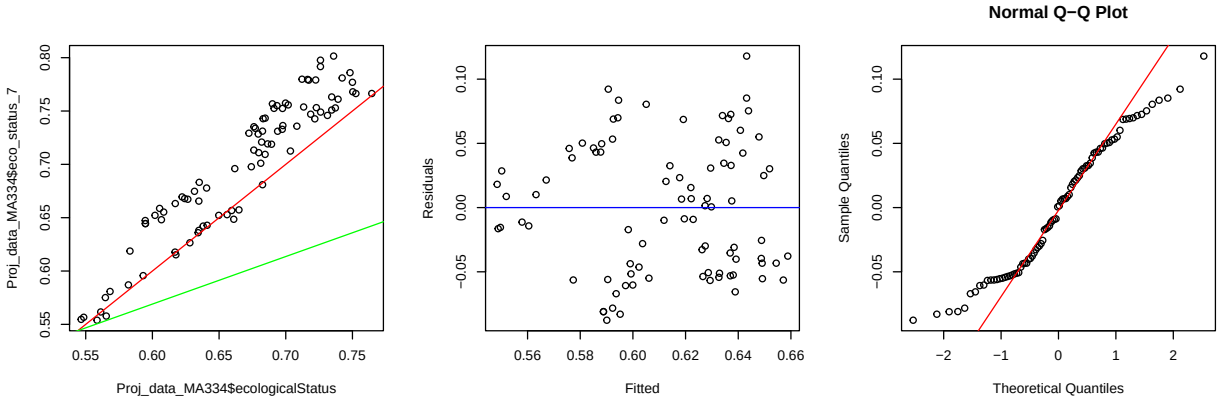
Kolmogorov-Smirnov (KS) test: This test can be used to determine whether two samples come from the same distribution. In this case, we can use a KS test to compare the distribution of biodiversity based on 7 taxonomic groups with the distribution based on 11 taxonomic groups.

	D	p_value
D	0.3181818	0.0002437

The result of the KS test shows that the D-value is 0.31818 with a p-value of 0.0002437. Since the p-value is much smaller than the significance level of 0.05, we can reject the null hypothesis that the two distributions are the same. Therefore, we can conclude that there is a significant difference in the biodiversity of the study area based on 7 taxonomic groups compared to 11 taxonomic groups.

Simple Linear Regression

Figure 1: Linear Regression



The simple linear regression results suggest a positive relationship between BD7 and BD11, with a slope coefficient of 1.11469, indicating that for each one-unit increase in BD11, we can expect an average increase of 1.11469 units in BD7.

When examining each period separately, we see a positive relationship between BD7 and BD11 for both Y70 and Y00. For Y70, the slope coefficient is 0.52272, indicating that for each one-unit increase in BD11, we can expect an average increase of 0.52272 units in BD7. For Y00, the slope coefficient is 1.16125945, indicating that for each one-unit increase in BD11, we can expect an average increase of 1.16125945 units in BD7.

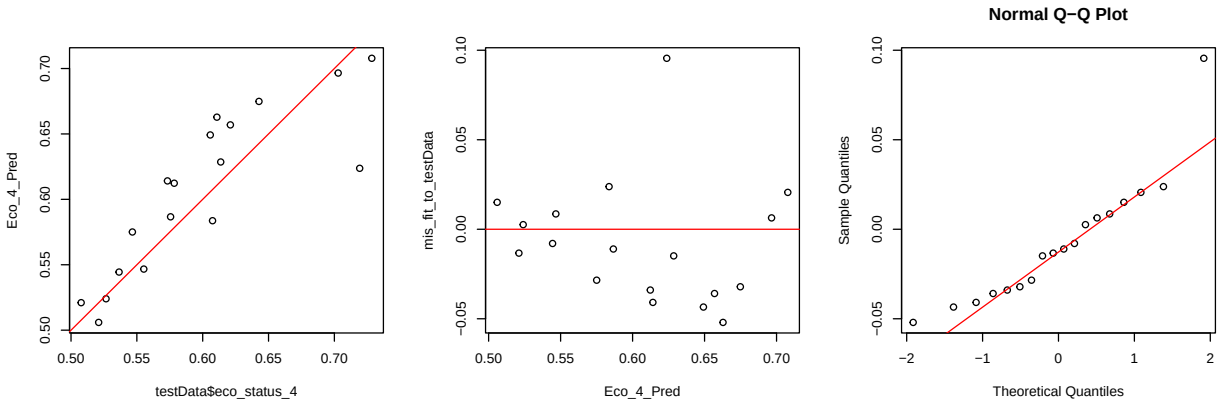
Overall, the results suggest that BD7 and BD11 are positively related, with a stronger relationship in the later period (Y00).

Multiple Linear Regression

The analysis aimed to assess the effect of seven ecological variables on a measure of biodiversity excluded from these seven variables.

First, the mean of the proportional species richness values for the remaining four taxonomic groups (BD4) was calculated by removing the seven variables from the eleven in the data set. Then, a multiple linear regression was performed with BD4 as the dependent variable and the seven ecological variables as the independent variables. Missing values in the mean_selected variable were checked and BD4 was added to the data set as a new variable.

Figure 2: Multiple Linear Regression



The relationship between BD4 and the mean of the seven ecological variables in Figure 2, shows a positive linear relationship. A linear regression analysis was then performed with BD4 as the dependent variable and the mean of the seven ecological variables as the independent variable. The results showed a significant positive effect of the seven ecological variables on BD4, with a slope of 0.44569, t-value of 5.355, p-value of 7.04e-07, and R-squared value of 0.25. The residuals were found to be normally distributed, and the model assumptions were met.

The model summary shows that only the variables Carabids and Vascular_plants had a significant effect on eco_status_4. The AIC value for the model was 75.11, which was lower than the AIC values for other models, indicating that this model was the best fit for the data. The multiple R-squared value for the model was 0.4242, indicating that the model explained 42.42% of the variance in eco_status_4.

The regression coefficients showed that the intercept was not significant. Vascular_plants had a positive effect on eco_status_4 (slope = 0.2140, t = 2.114, p-value = 0.038), while Carabids had a negative effect (slope = -0.2592, t = -2.439, p-value = 0.017). The other variables did not have a significant effect on eco_status_4.

The results of this analysis indicate that only two ecological variables, Carabids and Vascular_plants, had a significant effect on a measure of biodiversity excluded from the other seven variables. Vascular_plants had a positive effect on eco_status_4, while Carabids had a negative effect. These results suggest that these two variables are important predictors of biodiversity and should be considered in biodiversity conservation efforts.